Statistica Sinica Preprint No: SS-13-173R4

Title	The independence process in conditional quantile location-scale models and an application to testing for monotonicity
Manuscript ID	SS-13-173R4
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.2013.173
Complete List of Authors	Melanie Birke Natalie Neumeyer and Stanislav Volgushev
Corresponding Author	Natalie Neumeyer
E-mail	neumeyer@math.uni-hamburg.de

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

The independence process in conditional quantile location-scale models and an application to testing for monotonicity

Melanie Birke, Natalie Neumeyer* and Stanislav Volgushev

University of Bayreuth, University of Hamburg and University of Toronto

January 24, 2017

Abstract

In this paper the nonparametric quantile regression model is considered in a location-scale context. The asymptotic properties of the empirical independence process based on covariates and estimated

*corresponding author; e-mail: neumeyer@math.uni-hamburg.de

doi:10.5705/ss.2013.173

residuals are investigated. In particular an asymptotic expansion and weak convergence to a Gaussian process are proved. The results can be applied to test for validity of the location-scale model, and they allow one to derive various specification tests in conditional quantile location-scale models. A test for monotonicity of the conditional quantile curve is investigated. For the test for validity of the location-scale model, as well as for the monotonicity test, smooth residual bootstrap versions of Kolmogorov-Smirnov and Cramér-von Mises type test statistics are suggested. We give proofs for bootstrap versions of the weak convergence results. The performance of the tests is demonstrated in a simulation study.

AMS Classification: 62G10, 62G08, 62G30

Keywords and Phrases: bootstrap, empirical independence process, Kolmogorov-Smirnov test, model test, monotone rearrangements, nonparametric quantile regression, residual processes, sequential empirical process

Short title: The independence process in quantile models

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

1 Introduction

Quantile regression was introduced by Koenker and Bassett (1978) as an extension of least squares methods focusing on the estimation of the conditional mean function. Due to such attractive features as robustness with respect to outliers and equivariance under monotonic transformations that are not shared by the mean regression, it has since become increasingly popular in such fields as medicine, economics and environment modelling (see Yu et al. (2003) or Koenker (2005)). An important feature of quantile regression is its great flexibility. While mean regression aims at modelling the average behaviour of a variable Y given a covariate $X = x$, quantile regression allows one to analyse the impact of X in different regions of the distribution of Y by estimating several quantile curves simultaneously. For example, Fitzenberger et al. (2008) demonstrate that the presence of certain structures in a company can have different effects on upper and lower wages. For a more detailed discussion, see Koenker (2005).

Here we prove a weak convergence result for the empirical independence process of covariates and estimated errors in a nonparametric location-scale conditional quantile model. In turn, this suggests a test for monotonicity of the conditional quantile curve. To the authors' best knowledge, this is the

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

first time that those problems have been treated for the general nonparametric quantile regression model.

The empirical independence process results from the distance between a joint empirical distribution function and the product of the marginal empirical distribution functions. It can be used to test for independence; see Hoeffding (1948), Blum et al. (1961), and van der Vaart and Wellner (1996). When applied to covariates X and estimators of error terms $\varepsilon = (Y - q(X))/s(X)$ it can be used to test for validity of a location-scale model $Y = q(X) + s(X)\varepsilon$ with X and ε independent. Here the conditional distribution of Y, given $X = x$, allows for a location-scale representation $P(Y \leq y \mid X = x) = F_\varepsilon((y - q(x))/s(x))$, where F_ε denotes the error distribution function. Einmahl and Van Keilegom (2008a) consider such tests for location-scale models in a general setting (mean regression, trimmed mean regression,...). But the assumptions made there rule out the quantile regression case. Validity of a location-scale model means that the covariates have influence on the trend and on the dispersion of the conditional distribution of Y, but otherwise do not affect the shape of the conditional distribution (such models are frequently used, see Shim et al. (2009) and Chen et al. (2005)). Should the test reject independence of covariates and errors, there is evidence that the influence of the covariates on the response goes beyond location and scale effects. Our results can easily be adapted to test the validity of location models $P(Y \leq y \mid X = x) = F_\varepsilon(y - q(x))$; see also Einmahl and Van Keilegom (2008b) and Neumeyer (2009b) in the mean regression context.

If there is economic, physical, or biological evidence that a quantile curve is monotone, one might check by a statistical test that such a feature is reasonable. In classical mean regression there are various methods for testing monotonicity; see Bowman et al. (1998), Gijbels et al. (2000), Hall and Heckman (2001), Goshal et al. (2000), Durot (2003), Baraud et al. (2003), Domínguez-Menchero et al. (2005), Birke and Dette (2007), Wang and Meyer (2011), and Birke and Neumeyer (2013). While most tests are conservative and lack power against alternatives with only a small deviation from monotonicity, the method proposed by Birke and Neumeyer (2013) has better power in some situations and can detect local alternatives of order $n^{-1/2}$. For monotone estimators of a quantile function, see e.g. Cryer et al. (1972) and Robertson and Wright (1973) for median regression and Casady and Cryer (1976) and Abrevaya (2005) for general quantile regression. Still, testing whether a given quantile curve is increasing (decreasing) has received little attention aside from Duembgen (2002).

The paper is organized as follows. In Section 2 we present the location-scale model, give necessary assumptions and define the estimators. In Section 3 we introduce the independence process, derive asymptotical results and construct a test for validity of the model. Bootstrap data generation and asymptotic results for a bootstrap version of the independence process are discussed as well. The results derived there are modified in Section 4 to construct a test for monotonicity of the quantile function. In Section 5 we present a small simulation study, and we conclude in Section 6. Proofs are given in the supplementary materials.

2 The location-scale model, estimators and assumptions

For some fixed $\tau \in (0,1)$, consider the nonparametric quantile regression model of location-scale type (see e.g. He (1997)),

$$Y_i \;=\; q_\tau(X_i) + s(X_i)\varepsilon_i, \qquad i = 1,\ldots,n, \tag{2.1}$$

where $q_\tau(x) = F_Y^{-1}(\tau|x)$ is the τ-th conditional quantile function, (X_i, Y_i), $i = 1,\ldots,n$, is a bivariate sample of i.i.d. observations, $F_Y(\cdot|x) = P(Y_i \leq \cdot|X_i = x)$ denotes the conditional distribution function of Y_i given $X_i = x$,

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

and $s(x)$ denotes the median of $|Y_i - q_\tau(X_i)|$, given $X_i = x$. We assume that ε_i and X_i are independent and, hence, that ε_i has τ-quantile zero and $|\varepsilon_i|$ has median one, since

$$\tau = P\Big(Y_i \leq q_\tau(X_i) \,\Big|\, X_i = x\Big) = P(\varepsilon_i \leq 0)$$
$$\frac{1}{2} = P\Big(|Y_i - q_\tau(X_i)| \leq s(X_i) \,\Big|\, X_i = x\Big) = P(|\varepsilon_i| \leq 1).$$

Denote by F_ε the distribution function of ε_i. Then for the conditional distribution we obtain a location-scale representation as $F_Y(y|x) = F_\varepsilon((y - q_\tau(x))/s(x))$, where F_ε as well as q_τ and s are unknown.

Consider the case $\tau = \frac{1}{2}$ for example. This is a median regression model, that allows for heteroscedasticity, in the sense that the conditional median absolute deviation $s(X_i)$ of Y_i, given X_i, may depend on the covariate X_i. Here the median absolute deviation of a random variable Z is $\mathrm{MAD}(Z) = \mathrm{median}(|Z - \mathrm{median}(Z)|)$; it is the typical measure of scale (or dispersion) when the median is used as location measure.

Remark 1 Assuming $|\varepsilon_i|$ to have median one is not restrictive: if the model $Y_i = q_\tau(X_i) + \tilde{s}(X_i)\eta_i$ with η_i i.i.d. and independent of X_i and some positive function $\tilde{s}$ holds, the model $Y_i = q_\tau(X_i) + s(X_i)\varepsilon_i$ with $s(X_i) := \tilde{s}(X_i)F_{|\eta|}^{-1}(1/2)$, $\varepsilon_i := \eta_i/F_{|\eta|}^{-1}(1/2)$ also holds, where $F_{|\eta|}$ denotes the distribution function of $|\eta_i|$; then, in particular, $P(|\varepsilon_i| \leq 1) = P(|\eta_i| \leq F_{|\eta|}^{-1}(1/2)) = 1/2$.

Several non-parametric quantile estimators have been proposed (see e.g. Yu and Jones (1997, 1998), Takeuchi et al. (2006), Dette and Volgushev (2008), among others). We follow the last-named authors who proposed non-crossing estimates of quantile curves using a simultaneous inversion and isotonization of an estimate of the conditional distribution function. Thus, let

$$\hat{F}_Y(y|x) := (\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{W}\mathbf{Y} \tag{2.2}$$

with

$$\mathbf{X} = \begin{pmatrix} 1 & (x-X_1) & \dots & (x-X_1)^p \\ \vdots & \vdots & \dots & \vdots \\ 1 & (x-X_n) & \dots & (x-X_n)^p \end{pmatrix}, \qquad \mathbf{Y} := \Big(\Omega\Big(\frac{y-Y_1}{d_n}\Big), \dots, \Omega\Big(\frac{y-Y_n}{d_n}\Big)\Big)^t$$

$$\mathbf{W} = \mathrm{Diag}\Big(K_{h_n,0}(x-X_1), ..., K_{h_n,0}(x-X_n)\Big),$$

denote a smoothed local polynomial estimate (of order $p \geq 2$) of the conditional distribution function $F_Y(y|x)$, where $\Omega(\cdot)$ is a smoothed version of the indicator function and we used $K_{h_n,k}(x) := K(x/h_n)(x/h_n)^k$. Here K is a nonnegative kernel and d_n, h_n are bandwidths converging to 0 with increasing sample size. The estimator $\hat{F}_Y(y|x)$ can be represented as a weighted average

$$\hat{F}_Y(y|x) = \sum_{i=1}^{n} W_i(x)\Omega\Big(\frac{y - Y_i}{d_n}\Big). \tag{2.3}$$

Following Dette and Volgushev (2008) we consider a strictly increasing distribution function $G : \mathbb{R} \to (0, 1)$, a nonnegative kernel κ, and a bandwidth b_n, and define the functional

$$H_{G,\kappa,\tau,b_n}(F) := \frac{1}{b_n}\int_0^1 \int_{-\infty}^{\tau} \kappa\Big(\frac{F(G^{-1}(u)) - v}{b_n}\Big) dv du.$$

It is intuitively clear that $H_{G,\kappa,\tau,b_n}(\hat{F}_Y(\cdot|x))$, where $\hat{F}_Y$ is the estimator of the conditional distribution function as in (2.2), is a consistent estimate of $H_{G,\kappa,\tau,b_n}(F_Y(\cdot|x))$. If $b_n \to 0$, this quantity can be approximated as

$$\begin{aligned} H_{G,\kappa,\tau,b_n}(F_Y(\cdot|x)) \;&\approx\; \int_{\mathbb{R}} I\{F_Y(y|x) \le \tau\} dG(y) \\ &= \int_0^1 I\{F_Y(G^{-1}(v)|x) \le \tau\} dv \;=\; G \circ F_Y^{-1}(\tau|x) \end{aligned}$$

and, as a consequence, an estimate of the conditional quantile function $q_\tau(x) = F_Y^{-1}(\tau|x)$ is

$$\hat{q}_\tau(x) := G^{-1}(H_{G,\kappa,\tau,b_n}(\hat{F}_Y(\cdot|x))).$$

The scale function s is the conditional median of the distribution of $|e_i|$, given the covariate X_i, where $e_i = Y_i - q_\tau(X_i) = s(X_i)\varepsilon_i$, $i = 1, \ldots, n$. Hence, we

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

apply the quantile-regression approach to $|\hat{e}_i| = |Y_i - \hat{q}_\tau(X_i)|$, $i = 1, \ldots, n$, and obtain the estimator

$$\hat{s}(x) = G_s^{-1}(H_{G_s,\kappa,1/2,b_n}(\hat{F}_{|e|}(\cdot|x))) \,. \tag{2.4}$$

Here $G_s : \mathbb{R} \to (0,1)$ is a strictly increasing distribution function and $\hat{F}_{|e|}(\cdot|x)$ denotes the estimator of the conditional distribution function,

$$\hat{F}_{|e|}(y|x) = \sum_{i=1}^{n} W_i(x) I\{|\hat{e}_i| \le y\}, \tag{2.5}$$

with the same weights W_i as in (2.3). We further use the notation $F_e(\cdot|x) = P(e_1 \le \cdot|X_1 = x)$.

We need some assumptions on the kernel functions and the functions G, G_s used in the construction of the estimators.

(**K1**) The function K is a symmetric, positive, Lipschitz-continuous density with support $[-1,1]$. Moreover, the matrix $\mathcal{M}(K)$ with entries

$$(\mathcal{M}(K))_{k,l} = \mu_{k+l-2}(K) := \int u^{k+l-2} K(u) du$$

is invertible.

(**K2**) The function K is two times continuously differentiable, $K^{(2)}$ is Lipschitz continuous and, for $m = 0, 1, 2$, the set $\{x | K^{(m)}(x) > 0\}$ is a union of finitely many intervals.

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

(**K3**) The function Ω has derivative ω with support $[-1, 1]$, is a kernel of order p_ω, and is two times continuously differentiable with uniformly bounded derivatives.

(**K4**) The function κ is a symmetric, uniformly bounded density, and has one Lipschitz-continuous derivative.

(**K5**) The function $G : \mathbb{R} \to [0, 1]$ is strictly increasing; it is two times continuously differentiable in a neighborhood of the set $Q := \{q_\tau(x)|x \in [0, 1]\}$, and its first derivative is uniformly bounded away from zero on Q.

(**K6**) The function $G_s : \mathbb{R} \to (0, 1)$ is strictly increasing; it is two times continuously differentiable in a neighborhood of the set $S := \{s(x)|x \in [0, 1]\}$, and its first derivative is uniformly bounded away from zero on S.

For the data-generating process, we need the following conditions.

(**A1**) $X_1, \ldots, X_n$ are independent and identically distributed with distribution function F_X and Lipschitz-continuous density f_X, with support $[0, 1]$, that is uniformly bounded away from zero and infinity.

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

(**A2**) The function s is uniformly bounded and $\inf_{x\in[0,1]} s(x) = c_s > 0$.

(**A3**) The partial derivatives $\partial_x^k \partial_y^l F_Y(y|x), \partial_x^k \partial_y^l F_e(y|x)$ exist and are continuous and uniformly bounded on $\mathbb{R} \times [0,1]$ for $k \vee l \leq 2$ or $k + l \leq d$ for some $d \geq 3$.

(**A4**) The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed with strictly increasing distribution function F_ε (independent of X_i) and a density f_ε that is positive everywhere and continuously differentiable, such that $\sup_{y\in\mathbb{R}} |y f_\varepsilon(y)| < \infty$ and $\sup_{y\in\mathbb{R}} |y^2 f'_\varepsilon(y)| < \infty$. The ε_i have τ-quantile zero and $|\varepsilon_1|$ has median one.

(**A5**) For some $\alpha > 0$, $\sup_{u,y} |y|^\alpha (F_Y(y|u) \wedge (1 - F_Y(y|u))) < \infty$.

We assume that the bandwidth parameters satisfy the following.

(**BW**) $\dfrac{\log n}{nh_n(h_n \wedge d_n)^4} = o(1), \quad \dfrac{\log n}{nh_n^2 b_n^2} = o(1), \quad d_n^{2(p_\omega \wedge d)} + h_n^{2((p+1)\wedge d)} + b_n^4 = o(n^{-1}),$

with p_ω from (K3), d from (A3), and p the order of the local polynomial estimator in (2.2).

Remark 2 Assumptions (**A1**) and (**A2**) are mild regularity assumptions on the data-generating process. Assumption (**A5**) places a mild condition on the

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

tails of the error distribution, and is satisfied even for distribution functions without finite first moments. Assumptions (**A3**) and (**A4**), by the Implicit Function Theorem, imply that $x \mapsto q_\tau(x)$ and $x \mapsto s(x)$ are two times continuously differentiable with uniformly bounded derivatives. This kind of condition is quite standard in the non-parametric estimation and testing literature. Due to the additional smoothing of $\hat{F}_Y(y|x)$ in the y-direction, we require more than the existence of just the second-order partial derivatives of $F_Y(y|x)$. As to bandwidths, observe that if, for example, $d = p_\omega = p = 3$ and we set $d_n = h_n = n^{-1/6-\beta}$ for some $\beta \in (0, 1/30)$, $b_n = h_n^{-1/4-\alpha}$ such that $\alpha + \beta \in (0, 1/12)$, condition (**BW**) holds. ■

3 The independence process, asymptotic results and testing for model validity

As estimators for the errors we build residuals

$$\hat{\varepsilon}_i = \frac{Y_i - \hat{q}_\tau(X_i)}{\hat{s}(X_i)}, \quad i = 1, \ldots, n. \tag{3.1}$$

To use $\hat{q}_\tau$ for building the residuals $\hat{e}_i$, we need $h_n \le X_i \le 1 - h_n$. The estimation of s requires us to again stay away from boundary points and we use the restriction $2h_n \le X_i \le 1 - 2h_n$.

For $y \in \mathbb{R}$, $t \in [2h_n, 1-2h_n]$, we let

$$\hat{F}_{X,\varepsilon,n}(t,y) \tag{3.2}$$
$$:= \sum_{i=1}^n I\{\hat{\varepsilon}_i \le y\} I\{2h_n < X_i \le t\} \frac{1}{\sum_{i=1}^n I\{2h_n < X_i \le 1-2h_n\}}$$
$$= \frac{1}{n}\sum_{i=1}^n I\{\hat{\varepsilon}_i \le y\} I\{2h_n < X_i \le t\} \frac{1}{\hat{F}_{X,n}(1-2h_n) - \hat{F}_{X,n}(2h_n)},$$

where $\hat{F}_{X,n}$ denotes the usual empirical distribution function of the covariates $X_1, \ldots, X_n$. The empirical independence process compares the joint empirical distribution with the product of the corresponding marginal distributions, and we take

$$S_n(t,y) = \sqrt{n}\Big(\hat{F}_{X,\varepsilon,n}(t,y) - \hat{F}_{X,\varepsilon,n}(1-2h_n,y)\hat{F}_{X,\varepsilon,n}(t,\infty)\Big) \tag{3.3}$$

for $y \in \mathbb{R}$, $t \in [2h_n, 1-2h_n]$, and $S_n(t,y) = 0$ for $y \in \mathbb{R}$, $t \in [0, 2h_n) \cup (1-2h_n, 1]$.

Theorem 1 *Under the location-scale model (2.1) and assumptions (**K1**)-(**K6**), (**A1**)-(**A5**), and (**BW**),*

$$S_n(t,y) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\varepsilon_i \le y\} - F_\varepsilon(y) - \phi(y)\Big(I\{\varepsilon_i \le 0\} - \tau\Big) - \psi(y)\Big(I\{|\varepsilon_i| \le 1\} - \frac{1}{2}\Big)\Big)$$
$$\times \Big(I\{X_i \le t\} - F_X(t)\Big) + o_P(1)$$

uniformly with respect to $t \in [0,1]$ and $y \in \mathbb{R}$, where

$$\phi(y) = \frac{f_\varepsilon(y)}{f_\varepsilon(0)}\Big(1 - y\frac{f_\varepsilon(1) - f_\varepsilon(-1)}{f_{|\varepsilon|}(1)}\Big), \quad \psi(y) = \frac{y f_\varepsilon(y)}{f_{|\varepsilon|}(1)}$$

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

and $f_{|\varepsilon|}(y) = (f_\varepsilon(y) + f_\varepsilon(-y))I_{[0,\infty)}(y)$ *is the density of* $|\varepsilon_1|$. *The process* S_n *converges weakly in* $\ell^\infty([0,1]\times\mathbb{R})$ *to a centered Gaussian process* S *with covariance*

$$\begin{aligned}
\mathrm{Cov}(S(s,y),S(t,z)) \;&=\; (F_X(s\wedge t) - F_X(s)F_X(t)) \\
\times\Big[&F_\varepsilon(y\wedge z) - F_\varepsilon(y)F_\varepsilon(z) + \phi(y)\phi(z)(\tau-\tau^2) + \frac{1}{4}\psi(y)\psi(z) \\
&- \phi(y)(F_\varepsilon(z\wedge 0) - F_\varepsilon(z)\tau) - \phi(z)(F_\varepsilon(y\wedge 0) - F_\varepsilon(y)\tau) \\
&- \psi(y)\Big((F_\varepsilon(z\wedge 1) - F_\varepsilon(-1))I\{z>-1\} - \frac{1}{2}F_\varepsilon(z)\Big) \\
&- \psi(z)\Big((F_\varepsilon(y\wedge 1) - F_\varepsilon(-1))I\{y>-1\} - \frac{1}{2}F_\varepsilon(y)\Big) \\
&+ (\phi(y)\psi(z) + \phi(z)\psi(y))\Big(F_\varepsilon(0) - F_\varepsilon(-1) - \frac{1}{2}\tau\Big)\Big].
\end{aligned}$$

The proof is given in the supplementary materials.

Remark 3 The result can easily be adapted for location models $Y_i = q_\tau(X_i) + \varepsilon_i$ with ε_i and X_i independent: we just set $\hat{s} \equiv 1$ in the definition of the estimators. The asymptotic covariance in Theorem 1 then simplifies as ϕ reduces to $\phi(y) = f_\varepsilon(y)/f_\varepsilon(0)$ and $\psi(y) \equiv 0$.

We now discuss the testing of the null hypothesis of independence of error ε_i and covariate X_i in model (2.1).

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

Remark 4 Assume X_i and ε_i are dependent, but the other assumptions of Theorem 1 are valid, where (**A4**) is replaced by

(**A4'**) The conditional error distribution function $F_\varepsilon(\cdot|x) = P(\varepsilon_i \leq \cdot|X_i = x)$ fulfills $F_\varepsilon(0|x) = \tau$ and $F_\varepsilon(1|x) - F_\varepsilon(-1|x) = \frac{1}{2}$ for all x. It is strictly increasing and differentiable with density $f_\varepsilon(\cdot|x)$ such that $\sup_{x,y} |y f_\varepsilon(y|x)| < \infty$.

Then one can show that $S_n(t,y)/n^{1/2}$ converges in probability to $P(\varepsilon_i \leq y, X_i \leq t) - F_\varepsilon(y)F_X(t)$, uniformly with respect to y and t.

Remark 5 If the location-scale model is valid for some τ-th quantile regression function it is valid for every α-th quantile regression function, $\alpha \in (0,1)$. This follows from $q_\alpha(x) = F_\varepsilon^{-1}(\alpha)s(x) + q_\tau(x)$, a consequence of the representation of the conditional distribution function $F_Y(y|x) = F_\varepsilon((y - q_\tau(x))/s(x))$ (compare Remark 1). A similar statement is true for general location and scale measures, see e. g. Van Keilegom (1998), Prop. 5.1. Thus for testing the validity of the location-scale model one can restrict oneself to the median case $\tau = 0.5$.

Remark 6 Einmahl and Van Keilegom (2008a) consider a process similar to S_n for general location and scale models that rules out the quantile case

$q(x) = F^{-1}(\tau|x)$.

To test for the validity of a location-scale model we reject the null hypothesis of independence of X_i and ε_i for large values of, e. g., the Kolmogorov-Smirnov statistic

$$K_n = \sup_{t\in[0,1],y\in\mathbb{R}} |S_n(t,y)|$$

or the Cramér-von Mises statistic

$$C_n = \int_{\mathbb{R}} \int_{[0,1]} S_n^2(t,y)\, \hat{F}_{X,n}(dt)\, \hat{F}_{\varepsilon,n}(dy),$$

where $\hat{F}_{\varepsilon,n}(\cdot) = \hat{F}_{X,\varepsilon,n}(1-2h_n,\cdot)$.

Corollary 1 *Under the assumptions of Theorem 1,*

$$K_n \xrightarrow{d} \sup_{t\in[0,1],y\in\mathbb{R}} |S(t,y)| = \sup_{x\in[0,1],y\in\mathbb{R}} |S(F_X^{-1}(x),y)|$$
$$C_n \xrightarrow{d} \int_{\mathbb{R}} \int_{[0,1]} S^2(t,y) F_X(dt) F_\varepsilon(dy) = \int_{\mathbb{R}} \int_{[0,1]} S^2(F_X^{-1}(x),y)\, dx\, F_\varepsilon(dy).$$

The proof is given in Section S1 of the supplementary materials. The asymptotic distributions of the test statistics are independent of the covariate distribution F_X, but depend in a complicated manner on the error distribution F_ε, so we suggest a bootstrap version of the test. If $\mathcal{Y}_n = \{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ is the original sample, generate bootstrap errors as

$\varepsilon_i^* = \tilde{\varepsilon}_i^* + \alpha_n Z_i$ $(i = 1, \ldots, n)$, where α_n denotes a positive smoothing parameter, $Z_1, \ldots, Z_n$ are independent standard normals (independent of $\mathcal{Y}_n$), and $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_n^*$ are randomly drawn with replacement from the set of residuals $\{\hat{\varepsilon}_j \mid j \in \{1, \ldots, n\}, X_j \in (2h_n, 1-2h_n]\}$. Conditional on the original sample $\mathcal{Y}_n$, $\varepsilon_1^*, \ldots, \varepsilon_n^*$ are i.i.d. with distribution function

$$\tilde{F}_\varepsilon(y) = \frac{\frac{1}{n}\sum_{i=1}^n \Phi\Big(\frac{y-\hat{\varepsilon}_i}{\alpha_n}\Big) I\{2h_n < X_i \le 1-2h_n\}}{\hat{F}_{X,n}(1-2h_n) - \hat{F}_{X,n}(2h_n)}, \tag{3.4}$$

where Φ denotes the standard normal distribution function. The bootstrap error's τ-quantile is not exactly zero, but vanishes asymptotically. We use a smooth distribution to generate new bootstrap errors, see Neumeyer (2009a).

Now we build new bootstrap observations,

$$Y_i^* \;=\; \hat{q}_\tau(X_i) + \hat{s}(X_i)\varepsilon_i^*, \quad i = 1, \ldots, n.$$

Let $\hat{q}_\tau^*$ and $\hat{s}^*$ denote the quantile regression and scale function estimator defined analogously to $\hat{q}_\tau$ and $\hat{s}$, but based on the bootstrap sample $(X_1, Y_1^*), \ldots, (X_n, Y_n^*)$. Analogously to (3.3) the bootstrap version of the independence process is

$$S_n^*(t,y) \;=\; \sqrt{n}\Big(\hat{F}^*_{X,\varepsilon,n}(t,y) - \hat{F}^*_{X,\varepsilon,n}(1-4h_n, y)\hat{F}^*_{X,\varepsilon,n}(t,\infty)\Big)$$

for $t \in [4h_n, 1-4h_n]$, $y \in \mathbb{R}$, and $S_n^*(t,y) = 0$ for $t \in [0, 4h_n) \cup (1-4h_n, 1]$,

$y \in \mathbb{R}$. Here, similar to (3.2),

$$\hat{F}^*_{X,\varepsilon,n}(t,y) \;=\; \frac{1}{n}\sum_{i=1}^n I\{\hat{\varepsilon}^*_i \le y\} I\{4h_n < X_i \le t\} \frac{1}{\hat{F}_{X,n}(1-4h_n) - \hat{F}_{X,n}(4h_n)},$$

with $\hat{\varepsilon}^*_i = (Y^*_i - \hat{q}^*_\tau(X_i))/\hat{s}^*(X_i)$, $i = 1, \ldots, n$.

To obtain conditional weak convergence we need some assumptions.

(**B1**) For some $\delta > 0$

$$\frac{nh_n^2\alpha_n^2}{\log h_n^{-1} \log n} \to \infty, \quad \frac{n\alpha_n h_n}{\log n} \to \infty, \quad \frac{h_n}{\log n} = O(\alpha_n^{8\delta/3}), \quad n\alpha_n^4 = o(1)$$

and there exists a $\lambda > 0$ such that

$$\frac{nh_n^{1+\frac{1}{\lambda}}\alpha_n^{2+\frac{2}{\lambda}}}{\log h_n^{-1}(\log n)^{1/\lambda}} \to \infty.$$

(**B2**) $E[|\varepsilon_1|^{\max(\upsilon,2\lambda)}] < \infty$ for some $\upsilon > 1 + 2/\delta$, and with δ and λ from assumption (**B1**).

Here, (**B2**) can be relaxed to $E[|\varepsilon_1|^{2\lambda}] < \infty$ if the process is only considered for $y \in [-c, c]$ for some $c > 0$ instead of for $y \in \mathbb{R}$.

Theorem 2 *Under (2.1), (**K1**)-(**K6**), (**A1**)-(**A5**), (**BW**), and (**B1**)-(**B2**) conditionally on $\mathcal{Y}_n$, the process S^*_n converges weakly in $\ell^\infty([0,1]\times\mathbb{R})$ to the Gaussian process S defined in Theorem 1, in probability.*

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

A proof is given in the supplementary materials.

Remark 7 The bootstrap version of the Kolmogorov-Smirnov test statistic is $K_n^* = \sup_{t,y} |S_n^*(t,y)|$, and with

$$P(K_n^* \geq k_{n,1-\alpha}^* \mid \mathcal{Y}_n) = 1 - \alpha,$$

we reject the location-scale model if $K_n = \sup_{t,y} |S_n(t,y)| \geq k_{n,1-\alpha}^*$. The test has asymptotic level α. If the location-scale model is not valid, $K_n \to \infty$ in probability, while $k_{n,1-\alpha}^*$ converges to a constant. Thus the power of the test converges to one. Similar reasoning applies for the Cramér-von Mises test.

Remark 8 Recently, Sun (2006) and Feng et al. (2011) proposed to use wild bootstrap in the setting of quantile regression. To follow the approach of the last-named authors, one would define $\varepsilon_i^* = v_i \hat{\varepsilon}_i$ such that $P^*(v_i \hat{\varepsilon}_i \leq 0 | X_i) = \tau$, e. g.

$$v_i = \pm 1 \text{ with probability } \begin{cases} \frac{1-\tau}{\tau} & \text{if } \hat{\varepsilon}_i \geq 0 \\ \frac{\tau}{1-\tau} & \text{if } \hat{\varepsilon}_i < 0. \end{cases}$$

However, then when calculating the conditional asymptotic covariance (following the proof in the supplementary material), instead of $\tilde{F}_\varepsilon(y)$ the following term appears

$$\frac{1}{n} \sum_{i=1}^{n} P(v_i \hat{\varepsilon}_i \leq y \mid \mathcal{Y}_n) \stackrel{n\to\infty}{\longrightarrow} (1-\tau)(F_\varepsilon(y) - F_\varepsilon(-y)) + \tau.$$

One obtains $F_\varepsilon(y)$ (needed to obtain the same covariance as in Theorem 1) only for $y = 0$ or for median regression ($\tau = 0.5$) with symmetric error distributions, but not in general. Hence, wild bootstrap cannot be applied in the general context of procedures using empirical processes in quantile regression.

Remark 9 Under (2.1) the result of Theorem 1 can be applied to test for more specific model assumptions (e. g. testing goodness-of fit of a parametric model for the quantile regression function). The general approach is to build residuals $\hat{\varepsilon}_{i,0}$ that only under H_0 consistently estimate the errors (e. g. using a parametric estimator for the conditional quantile function). Recall the definition of $\hat{F}_{X,\varepsilon,n}$ in (3.2) and define analgously $\hat{F}_{X,\varepsilon_0,n}$ by using the residuals $\hat{\varepsilon}_{i,0}$. Then, analogously to (3.3), define

$$S_{n,0}(t,y) \;=\; \sqrt{n}\Big(\hat{F}_{X,\varepsilon_0,n}(t,y) - \hat{F}_{X,\varepsilon,n}(1-2h_n,y)\hat{F}_{X,\varepsilon,n}(t,\infty)\Big)$$

for $y \in \mathbb{R}$, $t \in [2h_n, 1-2h_n]$, and $S_{n,0}(t,y) = 0$ for $y \in \mathbb{R}$, $t \in [0,2h_n) \cup (1-2h_n, 1]$. With this process the discrepancy from the null hypothesis can be measured. This approach is considered in detail for the problem of testing monotonicity of conditional quantile functions in the next section. A related approach, which however does not assume the location-scale model, is suggested to test for significance of covariables in quantile regression models by Volgushev et al. (2013).

4 Testing for monotonicity of conditional quantile curves

In this section, we consider a test of the hypothesis H_0 that $q_\tau(x)$ is increasing in x. To this end we define an increasing estimator $\hat{q}_{\tau,I}$, which consistently estimates q_τ if the hypothesis H_0 is valid, and consistently estimates some increasing function $q_{\tau,I} \neq q_\tau$ under the alternative that q_τ is not increasing. For any function $g : [0,1] \to \mathbb{R}$ define its increasing rearrangement on $[a,b] \subset [0,1]$ as

$$\Gamma(g)(x) \;=\; \inf\left\{ z \in \mathbb{R} \,\Big|\, a + \int_a^b I\{g(t) \leq z\}\, dt \geq x \right\},\; x \in [a,b].$$

If g is increasing, then $\Gamma(g) = g|_{[a,b]}$. See Anevski and Fougères (2007) and Neumeyer (2007) who consider increasing rearrangements of curve estimators in order to obtain monotone versions of unconstrained estimators. We take Γ_n as Γ with $[a,b] = [h_n, 1-h_n]$, and define the increasing estimator $\hat{q}_{\tau,I} = \Gamma_n(\hat{q}_\tau)$, where $\hat{q}_\tau$ denotes the unconstrained estimator of q_τ in Section 2. Then $\hat{q}_{\tau,I}$ estimates the increasing rearrangement $q_{\tau,I} = \Gamma(q_\tau)$ of q_τ (with

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

$[a,b]=[0,1]$), and only under the hypothesis H_0 of an increasing regression function do we have $q_\tau = q_{\tau,I}$. In Figure 1 (right part) a non-increasing function q_τ and its increasing rearrangement $q_{\tau,I}$ are displayed.

We build (pseudo-) residuals

$$\hat{\varepsilon}_{i,I} = \frac{Y_i - \hat{q}_{\tau,I}(X_i)}{\hat{s}(X_i)} \tag{4.1}$$

to estimate the pseudo-errors $\varepsilon_{i,I} = (Y_i - q_{\tau,I}(X_i))/s(X_i)$ that coincide with the true errors $\varepsilon_i = (Y_i - q_\tau(X_i))/s(X_i)$ $(i = 1, \ldots, n)$ in general only under H_0. Note that we use $\hat{s}$ from (2.4) for the standardization and not an estimator built from the constrained residuals. If the true function q_τ is not increasing (e.g. as in Figure 1) and we calculate the pseudo-errors from $q_{\tau,I}$, they are no longer identically distributed. This effect is demonstrated in Figure 2 for a $\tau = 0.25$-quantile curve. To detect such discrepancies from the null hypothesis, we estimate the pseudo-error distribution for the covariate values $X_i \le t$ and compare with what is expected under H_0. Define $\hat{F}_{X,\varepsilon_I,n}$ analogously to (3.2), but using the constrained residuals $\hat{\varepsilon}_{i,I}$, $i = 1, \ldots, n$, and take

$$S_{n,I}(t,y) \;=\; \sqrt{n}\Big(\hat{F}_{X,\varepsilon_I,n}(t,y) - \hat{F}_{X,\varepsilon,n}(1-2h_n,y)\hat{F}_{X,\varepsilon,n}(t,\infty)\Big) \tag{4.2}$$

for $y \in \mathbb{R}$, $t \in [2h_n, 1-2h_n]$, and $S_{n,I}(t,y) = 0$ for $y \in \mathbb{R}$, $t \in [0, 2h_n) \cup (1-$

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

$2h_n, 1]$. For each fixed $t \in [0,1]$, $y \in \mathbb{R}$, for $h_n \to 0$ the statistic $n^{-1/2} S_{n,I}(t,y)$ consistently estimates the expectation

$$\begin{aligned} & E[I\{\varepsilon_{i,I} < y\} I\{X_i \le t\}] - F_\varepsilon(y) F_X(t) \\ = \; & E\Big[I\Big\{ \varepsilon_i < y + \frac{(q_{\tau,I} - q_\tau)(X_i)}{s(X_i)} \Big\} I\{X_i \le t\} \Big] - F_\varepsilon(y) F_X(t). \end{aligned}$$

For $K_n = \sup_{y \in \mathbb{R}, t \in [0,1]} |S_{n,I}(t,y)|$, $n^{-1/2} K_n$ estimates

$$K = \sup_{t \in [0,1], y \in \mathbb{R}} \left| \int_0^t \Big(F_\varepsilon\Big(y + \frac{(q_{\tau,I} - q_\tau)(x)}{s(x)} \Big) - F_\varepsilon(y) \Big) f_X(x)\, dx \right|.$$

Under $H_0 : q_{\tau,I} = q_\tau$ we have $K = 0$, and if $K = 0$, then also

$$\sup_{t \in [0,1]} \left| \int_0^t \Big(F_\varepsilon\Big(\frac{(q_{\tau,I} - q_\tau)(x)}{s(x)} \Big) - F_\varepsilon(0) \Big) f_X(x)\, dx \right| \; = \; 0.$$

Then it follows that $q_{\tau,I} = q_\tau$ is valid F_X-a. s. by the strict monotonicity of F_ε. Under the alternative we have $K > 0$ and K_n converges to infinity. If c is the $(1-\alpha)$-quantile of the distribution of $\sup_{t \in [0,1], y \in \mathbb{R}} |S(t,y)|$ with S from Theorem 1, the test that rejects H_0 for $K_n > c$ is consistent by the above argumentation and has asymptotic level α by the next theorem and the Continuous Mapping Theorem.

Theorem 3 *Under (2.1) and (**K1**)-(**K6**), (**A1**)-(**A5**), and (**BW**), under the null hypothesis* H_0 *and the assumption* $\inf_{x \in [0,1]} q_\tau'(x) > 0$ *the process*

$S_{n,I}$ converges weakly in $\ell^\infty([0,1]\times\mathbb{R})$ to the Gaussian process S defined in Theorem 1.

The proof is given in the supplementary materials.

Remark 10 We use the non-smooth monotone rearrangement estimators $\hat{q}_{\tau,I}$, while Dette et al. (2006) and Birke and Dette (2008) consider smooth versions of the increasing rearrangements in the context of monotone mean regression. Under suitable assumptions on the kernel k and bandwidths b_n the same weak convergence as in Theorem 3 holds for $S_{n,I}$ based on their approach.

For testing monotonicity we suggest a bootstrap version of the test as in Section 3, but applying the increasing estimator to build new observations as $Y_i^* = \hat{q}_{\tau,I}(X_i) + \hat{s}(X_i)\varepsilon_i^*$, $i = 1,\ldots,n$.

Theorem 4 *Under the assumptions of Theorem 3 and (**B1**)–(**B2**), the process $S^*_{n,I}$, conditionally on $\mathcal{Y}_n$, converges weakly in $\ell^\infty([0,1]\times\mathbb{R})$ to the Gaussian process S defined in Theorem 1, in probability.*

The proof is given in the supplementary materials. A consistent asymptotic level-α test is constructed as in Remark 7.

Remark 11 In the context of testing for monotonicity of mean regression curves Birke and Neumeyer (2013) based their tests on the observation that too many of the pseudo-errors are positive (see solid lines in Figure 2) on some subintervals of $[0,1]$ and too many are negative (see dashed lines) on other subintervals. Transferring this idea to the quantile regression model, one would consider a stochastic process

$$\tilde{S}_n(t,0) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(I\{\hat{\varepsilon}_{i,I}\le 0\}I\{2h_n<X_i\le t\}-\hat{F}_{X,\varepsilon,n}(1-2h_n,0)I\{2h_n<X_i\le t\}\Big)$$

or alternatively (because $\hat{F}_{X,\varepsilon,n}(1-2h_n,0)$ estimates the known $F_\varepsilon(0)=\tau$)

$$R_n(t) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(I\{\hat{\varepsilon}_{i,I}\le 0\}I\{X_i\le t\}-\tau I\{X_i\le t\}\Big)$$

where $t\in[0,1]$. For every $t\in[2h_n,1-2h_n]$ the processes count how many pseudo-residuals are positive up to covariates $\le t$. This term is then centered with respect to the estimated expectation under H_0 and scaled with $n^{-1/2}$. However, as can be seen from Theorem 3 the limit is degenerate for $y=0$, and hence we have under H_0 that

$$\sup_t|\tilde{S}_n(t,0)| = o_P(1). \tag{4.3}$$

Also, $\sup_{t\in[0,1]}|R_n(t)| = o_P(1)$ can be shown analogously. Hence, no critical values can be obtained for the Kolmogorov-Smirnov test statistics, and

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

those test statistics are not suitable for our testing purpose. To explain the negligibility (4.3) heuristically, consider the case $t = 1$ (now ignoring the truncation of covariates for simplicity of explanation). Then, under H_0, $n^{-1}\sum_{i=1}^n I\{\hat\varepsilon_{i,I} \le 0\}$ estimates $F_\varepsilon(0) = \tau$. But the information that ε_i has τ-quantile zero was already applied to estimate the τ-quantile function q_τ. Hence, one obtains $n^{-1}\sum_{i=1}^n I\{\hat\varepsilon_{i,I} \le 0\} - \tau = o_P(n^{-1/2})$. This observation is in accordance to the fact that $n^{-1}\sum_{i=1}^n \hat\varepsilon_i = o_P(n^{-1/2})$, when residuals are built from a mean regression model with centered errors (see Müller et al. (2004) and Kiwitt et al. (2008)).

Finally, consider the process

$$\begin{aligned}\tilde S_n(1-2h_n, y) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(& I\{\hat\varepsilon_{i,I} \le y\} I\{2h_n < X_i \le 1-2h_n\} \\ & - \hat F_{X,\varepsilon,n}(1-2h_n, y) I\{2h_n < X_i \le 1-2h_n\}\Big)\end{aligned}$$

i. e. the difference between the estimated distribution functions of pseudo-residuals $\hat\varepsilon_{i,I}$ and unconstrained residuals $\hat\varepsilon_i$ ($i = 1, \ldots, n$), respectively, scaled with $n^{1/2}$. An analogous process has been considered by Van Keilegom et al. (2008) for testing for parametric classes of mean regression functions. However, as can be seen from Theorem 3, in our case of testing for monotonicity the limit again is degenerate, i. e. $\mathrm{Var}(S(1, y)) = 0$ for all y, and hence $\sup_{y\in\mathbb{R}} |\tilde S_n(1, y)| = o_P(1)$. Similar observations can be made when typ-

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

ical distance based tests from lack-of-fit literature (for instance L^2-tests or residual process based procedures by Härdle and Mammen (1993) and Stute (1997), respectively) are considered in the problem of testing monotonicity of regression function, see Birke and Neumeyer (2013). The reason is that under H_0 the unconstrained and constrained estimators, $\hat{q}_\tau$ and $\hat{q}_{\tau,I}$, typically are first order asymptotically equivalent. This for estimation purposes very desirable property limits the possibilities to apply the estimator $\hat{q}_{\tau,I}$ for hypotheses testing.

5 Simulation results

In this section we show some simulation results for the bootstrap-based tests introduced in this paper. When available, we compare the results to already existing methods. Throughout, we choose the bandwidths according to condition (**BW**) as $d_n = 2(\hat{\sigma}^2/n)^{1/7}$, $h_n = (\hat{\sigma}^2/n)^{1/7}$, $b_n = \hat{\sigma}^2(1/n)^{2/7}$, and $\hat{\sigma}^2$ is the difference estimator proposed in Rice (1984) (see Yu and Jones (1997) for a related approach). The degree of the local polynomial estimators of location and scale (see equation (2.2)) was chosen to be three, the kernel K is the Gauss Kernel, while κ was chosen to be the Epanech-

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

nikov kernel. The function Ω was defined through $\Omega(t) = \int_{-\infty}^{t} \omega(x)dx$ where $\omega(x) := (15/32)(3 - 10x^2 + 7x^4)I\{|x| \leq 1\}$, a kernel of order 4 (see Gasser et al. (1985)). For the choices of G and G_s, we followed the procedure described in Dette and Volgushev (2008) who suggested a normal distribution such that the 5% and 95% quantiles coincide with the corresponding empirical quantities of the sample $Y_1, ..., Y_n$. The parameter α_n for generating the bootstrap residuals was chosen as $\alpha_n = 0.1n^{-1/4}\sqrt{2}\,\mathrm{median}(|\hat{\varepsilon}_1|, ..., |\hat{\varepsilon}_n|)$. The results under H_0 are based on 1000 simulation runs and 200 bootstrap replications, while the results under alternatives are based on 500 simulation runs and 200 bootstrap replications.

5.1 Testing for location and location-scale models

Testing the validity of location and location-scale models has previously been considered by Einmahl and Van Keilegom (2008a) and Neumeyer (2009b); we compare the properties of our test statistic with theirs. In testing the validity of location models (see Remark 3), we considered the data generation processes

$$\text{(model 1)} \quad Y|X = x \sim (x - 0.5x^2) + \frac{(1+ax)^{1/2}}{10}\mathcal{N}(0,1), \quad X \sim U[0,1],$$

$$\text{(model 2a)} \quad Y|X = x \sim (x - 0.5x^2) + \frac{1}{10}\Big(1 - \frac{1}{2c}\Big)^{1/2} t_c, \quad X \sim U[0,1],$$

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

$$\text{(model 2b)} \quad Y|X = x \sim (x - 0.5x^2) + \frac{1}{10}\Big(1 - (cx)^{1/4}\Big)^{1/2} t_{2/(cx)^{1/4}},$$

$$X \sim U[0,1],$$

$$\text{(model 3)} \quad Y|X = x, U = u \sim (x - 0.5x^2) + \Big(U - 0.5 - \frac{b}{6}(2x-1)\Big),$$

$$(X,U) \sim C(b).$$

Model 1 with parameter $a = 0$, model 2a with arbitrary parameter c, and model 3 with parameter $b = 0$ correspond to a location model, while models 1, 2b, and 3 with parameters $a, b, c \neq 0$ describe models that are not of this type. Here t_c denotes a $t-$distribution with c degrees of freedom (c not necessarily integer). Models 1 and 2b have been considered by Einmahl and Van Keilegom (2008a). Model 3 is from Neumeyer (2009b), with $(X,U) \sim C(b)$ generated as follows. Let X, V, W be independent $U[0,1]$-distributed random variables and take $U = \min(V, W/(b(1-2X)))$ if $X \leq \frac{1}{2}$, and $U = \max(V, 1 + W/(b(1-2X)))$ otherwise. Note that this data generation produces observations from the Farlie-Gumbel-Morgenstern copula if the parameter b is between -1 and 1.

Simulation results under the null are summarized in Table 1. The Kolmogorov-Smirnov (KS) and the Cramér-von Mises (CvM) bootstrap versions of the test hold the level quite well in all models considered, both for $n = 100$ and

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

$n = 200$ observations.

We looked at the power properties of the tests in models 1, 2a, and 3. The rejection probabilities are reported in Table 2, Table 3, and Table 4, respectively. For comparison, we include the results reported in Neumeyer (2009b) (noted N in the tables) and Einmahl and Van Keilegom (2008a) (noted EVK in the tables), where available. Neumeyer (2009b) considered several bandwidth parameters, while Einmahl and Van Keilegom (2008a) considered various types of test statistics (KS, CvM, and Anderson-Darling) and two types of tests (difference and estimated residuals). We have included the *best* values of the possible tests in Neumeyer (2009b) and Einmahl and Van Keilegom (2008a).

The tests of Neumeyer (2009b) and Einmahl and Van Keilegom (2008a) perform better for normal errors (Table 2), while our test seems to perform better for t errors (Table 3). This corresponds to intuition since for normal errors the mean provides an optimal estimator of location, while for heavier tailed distributions the median has an advantage. In almost all cases the CvM test outperforms the KS test. In model 3, the test of Neumeyer (2009b) performs better than our tests, with significantly higher power for $b = 1, 2$ and $n = 200$. The CvM version again has somewhat higher power than the

KS version of the test. Overall, we can conclude that the newly proposed testing procedures are competitive and can be particularly recommended for error distributions with heavier tails. In this, the CvM test seems to be preferable.

Please insert Tables 1, 2, 3 and 4 here

To evaluate the test for location-scale models, we considered the settings

$$\text{(model } 1_h) \quad Y|X = x \sim (x - 0.5x^2) + \frac{2+x}{10}\mathcal{N}(0,1), \quad X \sim U[0,1],$$

$$\text{(model } 2a_h) \quad Y|X = x \sim (x - 0.5x^2) + \frac{2+x}{10}\Big(1 - \frac{1}{2c}\Big)^{1/2} t_c, \quad X \sim U[0,1],$$

$$\text{(model } 2b_h) \quad Y|X = x \sim (x - 0.5x^2) + \frac{2+x}{10}\Big(1 - (cx)^{1/4}\Big)^{1/2} t_{2/(cx)^{1/4}}, \quad X \sim U[0,1],$$

$$\text{(model } 3_h) \quad Y|(X,U) = (x,u) \sim (x - 0.5x^2) + \frac{2+x}{10}\Big(U - 0.5 - b(2x-1)\Big) \quad (X,U) \sim C(b),$$

Models 1_h and $2b_h$ have been considered in Einmahl and Van Keilegom (2008a), while model 3_h is from Neumeyer (2009b). Simulation results corresponding to the different null models are collected in Table 5. In all three models the KS and the CvM test hold their level quite well for all sample

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

sizes, with both tests being slightly conservative for $n = 50$, and in model 1_h.

The power against alternatives in model $2b_h$ and 3_h is shown in Tables 6 and 7, respectively. In Table 7, the CvM version of the proposed test has higher (sometimes significantly so) power than the test of Neumeyer (2009b). One note is that the power of Neumeyer's test decreases for large values of b, while the power of our test continues to increase. This might be explained by the fact that for larger values of b, the variance of the residuals is extremely small, which probably leads to an instability of variance estimation.

In Table 6, the situation differs from the results in the homoscedastic model $2b$. In this setting, the tests proposed in this paper have no power for $n = 50$, or $n = 100$, even for the most extreme setting $b = 1$. The test of Einmahl and Van Keilegom (2008a) has less power than in the homoscedastic case, but is still able to detect that this model corresponds to the alternative. An intuitive explanation is that their residuals have the same variances while our residuals are scaled to have the same median absolute deviation. Under various alternative distributions, this leads to different power curves for the location-scale test. This difference is particularly extreme in the case of t-distributions. To illustrate this fact, recall in models which

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

are not of location-scale structure, $n^{-1/2}S_n(t,y)$ converges in probability to $P(\varepsilon_i^{ad} \leq y, X_i \leq t) - F_{\varepsilon^{ad}}(y)F_X(t)$, see Remark 4. Here, the residuals ε_i^{ad} are defined as $(Y_i - F_Y^{-1}(\tau|X_i))/s(X_i)$ with $s(x)$ denoting the conditional median absolute deviation of $Y_i - F_Y^{-1}(\tau|X_i)$ given $X_i = x$. A similar result holds for the residuals in EVK which take the form $\varepsilon_i^\sigma := (Y_i - m(X_i))/\sigma(X_i)$ where σ^2 denotes the conditional variance. One thus might expect that computing the quantities $K^{ad} := \sup_{t,y} |P(\varepsilon_i^{ad} \leq y, X_i \leq t) - F_{\varepsilon^{ad}}(y)F_X(t)|$ and $K^\sigma := \sup_{t,y} |P(\varepsilon_i^\sigma \leq y, X_i \leq t) - F_{\varepsilon^\sigma}(y)F_X(t)|$ will give some insights into the power properties of the KS test for residuals that are scaled in different ways. Indeed, numerical computations show that $K^\sigma/K^{ad} \approx 4.5$ which explains the large difference in power (note that the power for EVK reported in Table 6 is in fact the power of their Anderson-Darling test, the power of the KS test in EVK is lower). For a corresponding version of the CvM distance the ratio is roughly ten. We suspect that using a different scaling for the residuals would improve the power of the test in this particular model. However, since the optimal scaling depends on the underlying distribution of the residuals which is typically unknown, it seems difficult to implement an optimal scaling in practice. We leave this interesting question to future research. We do not present simulation results for the models with χ^2-distributed errors considered by Einmahl and Van Keilegom (2008a) and Neumeyer (2009b) for power simulations. For error distributions that are χ^2_b with $b < 2$, tests based on residuals do not hold their level, and the power characteristics described in those papers are a consequence of this fact. Weak convergence of the residual process requires errors to have a uniformly bounded density, which is not the case for chi-square distributions with degrees of freedom less than two.

Please insert Tables 5, 6 and 7 here

5.2 Testing for monotonicity of quantile curves in a location-scale setting

Please insert Figure 3 here

We considered the test for monotonicity of quantile curves that is introduced in Section 4. We simulated two models of location-scale type,

$$\text{(model 4)} \qquad Y|X = x \sim 1 + x - \beta e^{-50(x-0.5)^2} + 0.2\mathcal{N}(0,1), \quad X \sim U[0,1]$$

$$\text{(model 5)} \qquad Y|X = x \sim \frac{x}{2} + 2(0.1 - (x-0.5)^2)\mathcal{N}(0,1), \quad X \sim U[0,1].$$

The results for models 4 and 5 are reported in Table 8 and Table 9, respectively. In model 4, all quantile curves are parallel and so all quantile curves

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

have a similar monotonicity behavior. In particular, the parameter value $\beta = 0$ corresponds to strictly increasing quantile curves, for $\beta = 0.15$ the curves have a flat spot, and for $\beta > 0.15$ the curves have a small decreasing bump that gets larger for larger values of β. The median curves for different values of β are depicted in Figure 3, and the 25% quantile curves are parallel to the median curves with exactly the same shape. We performed the tests for two different quantile curves ($\tau = 0.25$ and $\tau = 0.5$) and in both cases the test has a slowly increasing power for increasing values of β and sample size. The case $\beta = 0.45$ is already recognized as alternative at $n = 50$, while for $\beta = 0.25$ the test only starts to show some power for $n = 200$.

In model 5, the median is a strictly increasing function while the outer quantile curves are not increasing. In Table 9, we report the simulation results for the quantile values $\tau = 0.25, \tau = 0.5$ and $\tau = 0.75$ and sample sizes $n = 50, 100, 200$. For $n = 50$, the observed rejection probabilities are slightly above the nominal critical values (for $\tau = 0.5$), and the cases $\tau = 0.25$ and $\tau = 0.75$ are recognized as alternatives. For $n = 100, 200$, the test holds its level for $\tau = 0.5$ and also shows a slow increase in power at the other quantiles. The increase is not really significant when going from $n = 50$ to $n = 100$ for $\tau = .25$, and not present for $\tau = .75$. For $n = 200$, the test

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

clearly has more power compared to $n = 50$. Overall, we can conclude that the proposed test shows a satisfactory behavior.

Please insert Tables 8 and 9 here

6 Conclusion

The paper at hand considered location-scale models in the context of nonparametric quantile regression. A test for model validity was investigated, based on the empirical independence process of covariates and residuals built from nonparametric estimators for the location and scale functions. The process converges weakly to a Gaussian process. A bootstrap version of the test was investigated in theory and by means of a simulation study. We considered in detail the testing for monotonicity of a conditional quantile function in theory as well as in simulations. Similarly other structural assumptions on the location or the scale function can be tested. A small simulation study demonstrated that the proposed method works well. All weak convergence results are proved in the supplementary materials.

Acknowledgments: We thank two anonymous referees and the associate editor for careful reading and for very constructive suggestions to improve the paper. Our special thanks go to one of the referees for several very careful readings of the manuscript and insightful comments. Part of this work was conducted while Stanislav Volgushev was postdoctoral fellow at the Ruhr University Bochum, Germany. During that time Stanislav Volgushev was supported by the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823), Teilprojekt (C1), of the Deutsche Forschungsgemeinschaft.

Supplementary materials: Proofs are presented in detail in the supplementary materials.

References

Abrevaya, J. (2005). Isotonic quantile regression: asymptotics and bootstrap. *Sankhyā* **67**, 187–199.

Anevski, D. and Fougères, A.-L. (2007). Limit properties of the monotone rearrangement for density and regression function estimation. arXiv:0710.4617v1

Baraud, Y. Huet, S. and Laurent, B. (2003). Adaptive tests of qualitative hypotheses. *ESAIM, Probab. Statist.* **7**, 147–159.

Birke, M. and Dette, H. (2007). Testing strict monotonicity in nonparametric regression. *Math. Meth. Statist.* **16**, 110–123.

Birke, M. and Dette, H. (2008). A note on estimating a smooth monotone regression by combining kernel and density estimates. *J. Nonparam. Statist.* **20**, 679–690.

Birke, M. and Neumeyer, N. (2013). Testing monotonicity of regression functions - an empirical process approach. *Scand. J. Statist.* **40**, 438–454.

Blum, J.R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution functions. *Ann. Math. Stat.* **32**, 485–498.

Bowman, A. W., Jones, M. C. and Gijbels, I. (1998). Testing monotonicity of regression. *J. Comput. Graph. Stat.* **7**, 489–500.

Casady, R.J. and Cryer, J.D. (1976). Monotone percentile regression. *Ann. Statist.* **4**, 532–541.

Chen, S., Dahl, G.B. and Khan, S. (2005). Nonparametric identification and estimation of a censored location-scale regression model. *J. Amer. Statist. Assoc.* **100**, 212–221.

Cryer, J.D., Robertson, T., Wright, F.T. and Casady, R.J. (1972). Monotone median regression. *Ann. Math. Statist.* **43**, 1459–1469.

Dette, H., Neumeyer, N. and Pilz, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* **12**, 469–490.

Dette, H. and Volgushev, S. (2008). Non–crossing nonparametric estimates of quantile curves. *J. Roy. Stat. Soc.* **B 70**, 609–627.

Domínguez-Menchero, J., González-Rodríguez, G. and López-Palomo, M. J. (2005). An L2 point of view in testing monotone regression *J. Nonparam. Statist.* **17**, 135–153.

Duembgen, L. (2002). Application of local rank tests to nonparametric regression. *J. Nonparametr. Stat.* **14**, 511–537.

Durot, C. (2003). A Kolmogorov-type test for monotonicity of regression. *tatist. Probab. Lett.* **63**, 425–433.

Einmahl, J.H.J. and Van Keilegom, I. (2008a). Specification tests in nonparametric regression. *J. Econometr.* **143**, 88–102.

Einmahl, J.H.J. and Van Keilegom, I. (2008b). Tests for independence in nonparametric regression. *Statist. Sinica* **18**, 601–616.

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

Fitzenberger, B., Kohn, K. and Lembcke, A. (2008). Union Density and Varieties of Coverage: The Anatomy of Union Wage Effects in Germany. IZA Working Paper No. 3356.
Available at SSRN: http://ssrn.com/abstract=1135932

Gasser, T., Müller, H. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Stat. Soc.* **B**, 238–252.

Gijbels, I., Hall, P., Jones, M. C. and Koch, I. (2000). Tests for monotonicity of a regression mean with guaranteed level. *Biometrika* **87**, 663–673.

Ghosal, S., Sen, A. and van der Vaart, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.* **28**, 1054–1082.

Hall, P. and Heckman, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28**, 20–39.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926–1947.

He, X. (1997). Quantile Curves without Crossing. *Am. Stat.* **51**, 186–192.

Hoeffding, W. (1948). A nonparametric test of independence. *Ann. Math. Statist.* **19**, 546–557.

Kiwitt, S., Nagel, E.-R. and Neumeyer, N. (2008). Empirical Likelihood Estimators for the Error Distribution in Nonparametric Regression Models. *Math. Meth. Statist.* **17**, 241–260.

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

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

Müller, U. U., Schick, A. and Wefelmeyer, W. (2004). Estimating linear functionals of the error distribution in nonparametric regression. *J. Statist. Plann. Inf.* **119**, 75–93.

Neumeyer, N. (2007). A note on uniform consistency of monotone function estimators. *Statist. Probab. Lett.* **77**, 693–703.

Neumeyer, N. (2009a). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scand. J. Statist.* **36**, 204–228.

Neumeyer, N. (2009b). Testing independence in nonparametric regression. *J. Multiv. Anal.* **100**, 1551–1566.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12**, 1215–1230.

Robertson, T. and Wright, F.T. (1973). Multiple isotonic median regression. *Ann. Statist.* **1**, 422–432.

Shim, J., Hwang, C. and Seok, H.K. (2009). Non-crossing quantile regression via doubly penalized kernel machine. *Comput. Statist.* **24**, 83–94.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613–641.

Sun, Y. (2006). A consistent nonparametric equality test of conditional quantile functions. *Econometric Th.* **22**, 614–632.

Takeuchi, I., Le, Q.V., Sears, T.D., and Smola, A.J. (2006). Nonparametric quantile regression. *Journal of Machine Learning Research*, **7**, 1231–1264.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer, New York.

Van Keilegom, I. (1998). *Nonparametric estimation of the conditional distribution in regression with censored data.* PhD thesis, University of Hasselt, Belgium. available at http.//www.ibiostat.be/publications/.

Van Keilegom, I., González–Manteiga, W. and Sánchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST* **17**, 401–415.

Volgushev, S., Birke, M., Dette, H. and Neumeyer, N. (2013). Significance testing in quantile regression. *Electron. J. Stat.* **7**, 105–145.

Wang, J.C. and Meyer, M.C. (2011). Testing the monotonicity or convexity of a function using regression splines. *Can. J. Stat.* **39**, 89–107.

Yu, K., and Jones, M.C. (1997). A comparison of local constant and local linear regression quantile estimators. *Comput. Stat. Data Anal.* **25**, 159–166.

Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228–237.

Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: applications and current research areas. *J. Roy. Stat. Soc. D (The Statistician)* **52**, 331–350.

MELANIE BIRKE, Universität Bayreuth, Fakultät für Mathematik, Physik und Informatik, 95440 Bayreuth, Germany, e-mail: Melanie.Birke@uni-bayreuth.de

NATALIE NEUMEYER, Universität Hamburg, Fachbereich Mathematik, Bundesstraße 55, 20146 Hamburg, Germany, e-mail: neumeyer@math.uni-hamburg.de

STANISLAV VOLGUSHEV, University of Toronto, Department of Statistical Sciences 100 St. George Street, Toronto, Ontario, Canada, e-mail: volgushe@utstat.toronto.edu

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

		model 1 $a = 0$	model 2a $c = 2$	model 3 $b = 0$
KS	$n = 100$	0.034	0.039	0.045
CvM	$n = 100$	0.029	0.044	0.053
KS	$n = 200$	0.034	0.039	0.050
CvM	$n = 200$	0.046	0.049	0.062

Table 1: *Rejection probabilities for testing the validity of a location model under various H_0 scenarios, the nominal level is $\alpha = 5\%$.*

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

	a	0	1	2.5	5	10
KS	$n = 100$	0.032	0.078	0.16	0.23	0.444
CvM	$n = 100$	0.038	0.128	0.364	0.568	0.746
N	$n = 100$	0.054	0.190	0.506	0.734	0.884
EVK	$n = 100$	0.072	0.132	0.316	0.524	0.668
KS	$n = 200$	0.034	0.144	0.292	0.586	0.784
CvM	$n = 200$	0.046	0.296	0.632	0.9	0.976
N	$n = 200$	0.044	0.390	0.860	0.976	0.972
EVK	$n = 200$	0.066	0.376	0.788	0.960	1.00

Table 2: *Rejection probabilities for testing the validity of a location model under the alternative in model 1 for different values of the parameter a, nominal level is $\alpha = 5\%$*

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

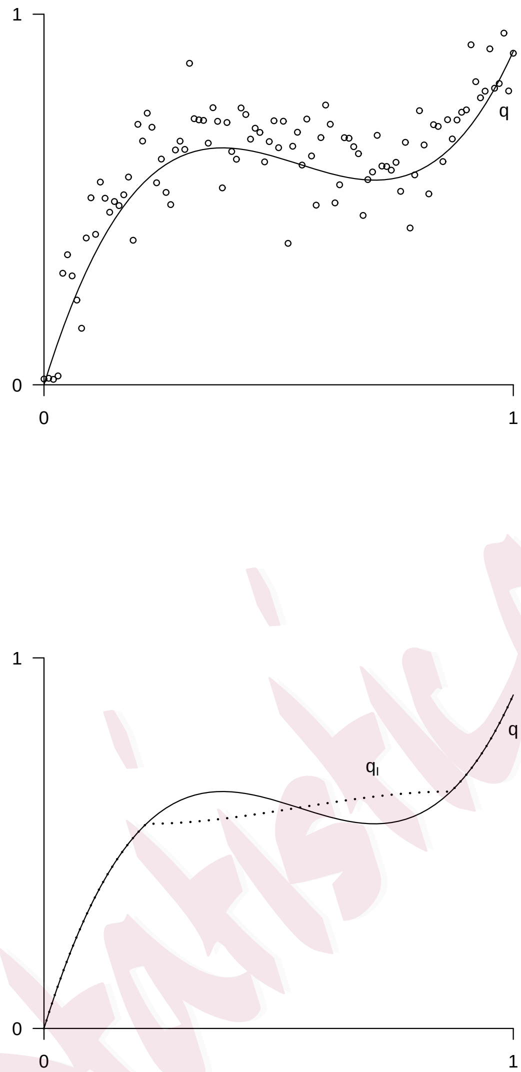

Figure 1: *Left part: True nonincreasing function q_τ for $\tau = 0.25$ with scatter-plot of a typical sample. Right part: q_τ (solid line) and increasing rearrangement $q_{\tau,I}$ (dotted line).*

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

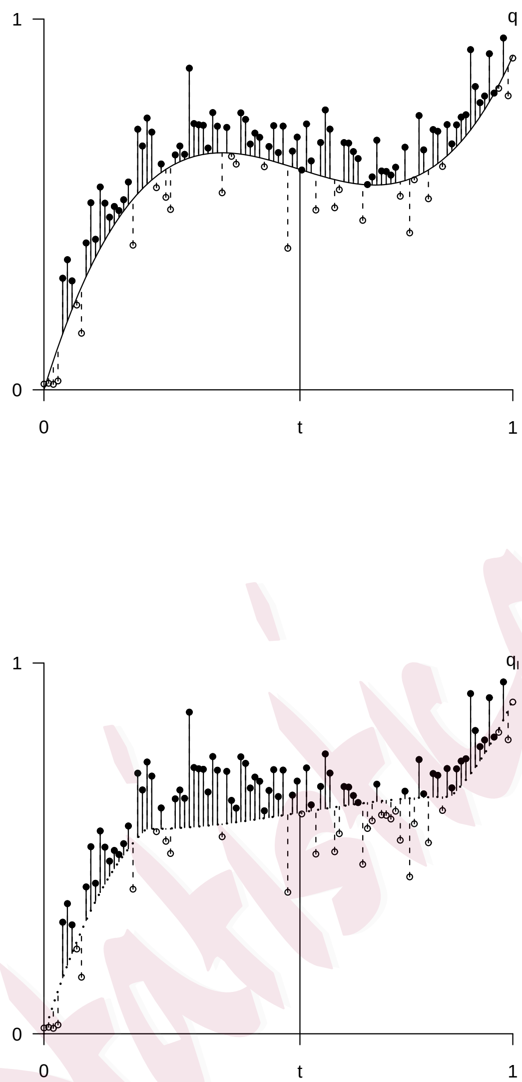

Figure 2: *Left part: True nonincreasing function q_τ for $\tau = 0.25$ and errors for the sample shown in Figure 1. Right part: Increasing rearrangement $q_{\tau,I}$ and pseudo-errors. (Positive errors are marked by solid points and solid lines, negative errors marked by circles and dashed lines.)*

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

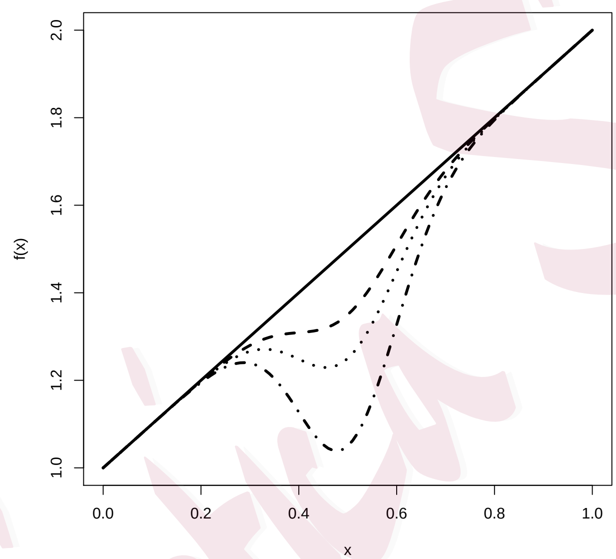

Figure 3: *The function* $x \mapsto 1 + x - \beta e^{-50(x-0.5)^2}$ *for values* $\beta = 0$ *(solid line),* $\beta = 0.15$ *(dashed line),* $\beta = 0.25$ *(dotted line),* $\beta = 0.45$ *(dash-dotted line), respectively. This is the median function in model 4, see Section 5.2.*

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

	c	.2	.4	.6	.8	1
KS	$n = 100$	0.044	0.074	0.120	0.194	0.390
CvM	$n = 100$	0.082	0.124	0.218	0.414	0.768
N	$n = 100$	0.096	0.120	0.224	0.420	0.676
EVK	$n = 100$	0.116	0.160	0.224	0.360	0.612
KS	$n = 200$	0.08	0.136	0.222	0.4	0.762
CvM	$n = 200$	0.118	0.29	0.49	0.792	0.996
N	$n = 200$	0.156	0.216	0.412	0.688	0.904
EVK	$n = 200$	0.124	0.216	0.344	0.584	0.944

Table 3: *Rejection probabilities for testing the validity of a location model under the alternative in model 2b for different values of the parameter c, nominal level is* $\alpha = 5\%$

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

	b	0	1	2	3	5
KS	$n = 100$	0.045	0.094	0.154	0.306	0.712
CvM	$n = 100$	0.053	0.128	0.240	0.576	0.968
N	$n = 100$	0.024	0.172	0.284	0.452	0.662
KS	$n = 200$	0.050	0.134	0.31	0.518	0.906
CvM	$n = 200$	0.062	0.254	0.538	0.92	1
N	$n = 200$	0.034	0.620	0.926	0.998	1.000

Table 4: *Rejection probabilities for testing the validity of a location model under the alternative in model 3 for different values of the parameter* b*, nominal level is* $\alpha = 5\%$

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

		model 1_h	model $2a_h$ $c = 2$	model 3_h $b = 0$
KS	$n = 50$	0.025	0.026	0.023
CvM	$n = 50$	0.022	0.026	0.034
KS	$n = 100$	0.031	0.037	0.037
CvM	$n = 100$	0.029	0.031	0.041
KS	$n = 200$	0.024	0.044	0.057
CvM	$n = 200$	0.028	0.044	0.062

Table 5: *Rejection probabilities for the testing the validity of a location-scale model under various H_0 scenarios, the nominal level is $\alpha = 5\%$.*

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

	c	1
KS	$n = 50$	0.032
CvM	$n = 50$	0.034
EVK	$n = 50$	0.262
KS	$n = 100$	0.046
CvM	$n = 100$	0.04
EVK	$n = 100$	0.478

Table 6: *Rejection probabilities for testing the validity of a location-scale model under the alternative in model* $2b_h$, *the nominal level is* $\alpha = 5\%$

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

	b	0	1	2	3	5
KS	$n = 100$	0.037	0.212	0.344	0.546	0.878
CvM	$n = 100$	0.041	0.368	0.658	0.922	0.992
N	$n = 100$	0.036	0.278	0.388	0.190	0.156
KS	$n = 200$	0.057	0.452	0.646	0.8	0.972
CvM	$n = 200$	0.062	0.802	0.966	1	1
N	$n = 200$	0.035	0.630	0.774	0.402	0.268

Table 7: *Rejection probabilities for testing the validity of a location-scale model under the alternative in model* 3_h *for different values of the parameter* b*, the nominal level is* $\alpha = 5\%$.

Statistica Sinica: Preprint
doi:10.5705/ss.2013.173

	$\tau = 0.25$			$\tau = 0.5$		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
$\beta = 0$	0.020	0.020	0.026	0.025	0.023	0.026
$\beta = 0.15$	0.024	0.027	0.050	0.027	0.047	0.060
$\beta = 0.25$	0.028	0.057	0.126	0.037	0.053	0.154
$\beta = 0.45$	0.140	0.202	0.410	0.084	0.154	0.344

Table 8: *Rejection probabilities for the test for monotonicity of quantile curves in model 4. The nominal level is* $\alpha = 5\%$.

	$n = 50$	$n = 100$	$n = 200$
$\tau = 0.25$	0.23	0.262	0.376
$\tau = 0.5$	0.073	0.061	0.043
$\tau = 0.75$	0.181	0.180	0.296

Table 9: *Rejection probabilities for the test for monotonicity of quantile curves in model 5. Different rows correspond to the* $0.25, 0.5$ *and* 0.75 *quantile curves, respectively. The nominal level is* $\alpha = 5\%$.