J. R. Statist. Soc. B (2008)
70, *Part* 3, *pp.* 609–627

Non-crossing non-parametric estimates of quantile curves

Holger Dette and Stanislav Volgushev

Ruhr-Universität Bochum, Germany

[Received May 2007. Final revision December 2007]

Summary. Since the introduction by Koenker and Bassett, quantile regression has become increasingly important in many applications. However, many non-parametric conditional quantile estimates yield crossing quantile curves (calculated for various $p \in (0,1)$). We propose a new non-parametric estimate of conditional quantiles that avoids this problem. The method uses an initial estimate of the conditional distribution function in the first step and solves the problem of inversion and monotonization with respect to $p \in (0,1)$ simultaneously. It is demonstrated that the new estimates are asymptotically normally distributed with the same asymptotic bias and variance as quantile estimates that are obtained by inversion of a locally constant or locally linear smoothed conditional distribution function. The performance of the new procedure is illustrated by means of a simulation study and some comparisons with the currently available procedures which are similar in spirit with the method proposed are presented.

Keywords: Conditional distribution; Crossing quantile curves; Local linear estimate; Nadaraya–Watson estimate; Quantile estimation

1. Introduction

Quantile regression was introduced by Koenker and Bassett (1978) as a supplement to least squares methods focusing on the estimation of the conditional mean function. The very general technique of estimating families of conditional quantile curves yields a great extension of parametric and non-parametric regression methods. Often the graphs of several conditional quantiles are used to represent the basic features of the entire conditional distribution of the response Y given the explanatory variable X. Applications of quantile regression include such important areas as medicine, economics and environment modelling. For more details we refer the interested reader to Yu *et al.* (2003) and Koenker (2005).

In this paper we consider the problem of non-parametric estimation of conditional quantile functions. Since the seminal work of Koenker and Bassett (1978), who discussed parametric quantile estimates, several researchers have proposed non-parametric methods in this context (see Chaudhuri (1991), Koenker *et al.* (1992, 1994) and Yu and Jones (1997, 1998) among many others). One very popular approach is to combine the concept of locally constant or locally linear estimation with the idea of using a 'check function' for the calculation of quantiles. More precisely, if $\{(X_i, Y_i)\}_{i=1}^n$ denotes a bivariate sample, the local constant (or Nadaraya–Watson) estimate of the pth conditional quantile of the conditional distribution $F(y|x)$ is defined as

$$\bar{q}_p(x) = \arg\min_a \left\{ \sum_{i=1}^n \rho_p(Y_i - a)\, K\left(\frac{x - X_i}{h} \right) \right\}, \tag{1.1}$$

Address for correspondence: Holger Dette, Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.
E-mail: holger.dette@rub.de

 1369–7412/08/70609

and the local linear estimate $\hat{q}_p(x)$ is obtained as $\hat{q}_p(x)=\hat{a}$, where $(\hat{a},\hat{b})$ minimizes the function

$$\sum_{i=1}^{n}\{\rho_p(Y_i-a-b(x-X_i))\}\,K\left(\frac{x-X_i}{h}\right) \tag{1.2}$$

(see for example Yu and Jones (1998) or Koenker (2005), chapter 7). In expressions (1.1) and (1.2)

$$\rho_p(z)=pz\,I_{[0,\infty)}(z)-(1-p)z\,I_{(-\infty,0)}(z)$$

denotes the check function, K is a kernel function and h a bandwidth tending to 0 with increasing sample size. As an alternative, several researchers have proposed to estimate the conditional distribution function non-parametrically,

$$\hat{F}(y|x)=\sum_{i=1}^{n}w_i(x)\,I(Y_i\leqslant y), \tag{1.3}$$

where $w_i(x)$ denotes a weight function depending on x and the predictors $\{X_i\}_{i=1}^n$, and then obtain the estimate of the quantile by solving the equation $p=\hat{F}(y|x)$. The solution of this equation is not necessarily unique (see Yu and Jones (1998)). Typical choices for the weights include the Nadaraya–Watson or local linear weights (see also Hall *et al.* (1999), who proposed adjusted Nadaraya–Watson weights).

It has been pointed out by several researchers (see for example He (1997), Yu *et al.* (2003) or Koenker (2005), chapter 7) that the estimates that are based on the concepts (1.1)–(1.3) have an important drawback in practical applications. For example local polynomial techniques—although very attractive from the viewpoint of mathematical efficiency—do not yield estimates respecting the logical monotonicity requirements with respect to the probability p. As a consequence, different quantile curves based on such estimators may cross, which is an undesirable feature in applications.

To demonstrate the logical problems arising in crossing estimates of quantile curves we consider the actual measurements of the relative change in bone mineral density at the spine in adolescents as a function of time (see Bachrach *et al.* (1999), Hastie *et al.* (2001) or Takeuchi *et al.* (2006)). In Fig. 1(a) we display the 10%, 20%, ..., 90% conditional quantile curves that are produced by the estimate of Koenker *et al.* (1994) for males and females. These estimates are computed by using the function `rqss` from the 'quantreg' package of Koenker (2007), where the smoothing parameter was chosen as $\lambda=5$. Fig. 1(b) shows the curves resulting from the method that is proposed in this paper. Although the overall picture that is provided by both estimation techniques is essentially the same, there is an a important difference: the quantile curves of Koenker *et al.* (1994) cross whereas the estimates that are proposed in this paper do not cross.

Several researchers have proposed procedures to avoid the embarrassment of quantile crossing. He (1997) suggested a restricted version of regression quantiles, whereas Yu and Jones (1998) suggested a double-kernel smoothing method and in the second step a minor modification of this estimate, such that the corresponding quantile curves are monotone as a function of the probability p. Hall *et al.* (1999) proposed an adjusted Nadaraya–Watson estimate which modifies the classical Nadaraya–Watson weights such that the resulting estimate of the conditional distribution function is monotone and asymptotically equivalent to the local linear method. The properties of positivity and monotonicity of this method allow us to obtain estimates of the conditional quantile from the conditional distribution estimate. However, it seems to be difficult to apply this methodology to local polynomial estimates of higher order. Note that this is straightforward for methods (1.1) and (1.2) that are based on the check function but at the cost of obtaining crossing quantile curve estimates; see Koenker (2005), chapter 7. Recently

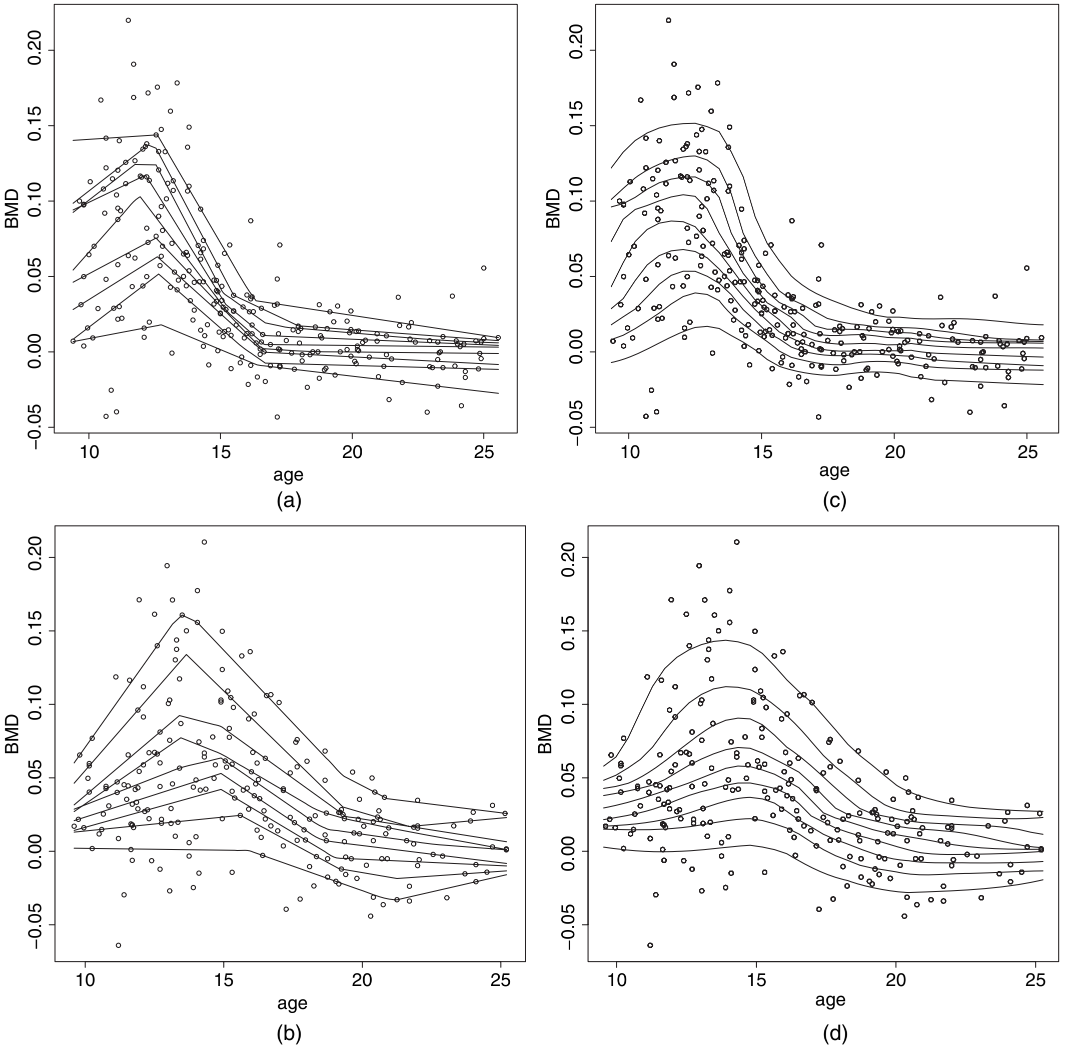

Fig. 1. (a), (b) Spline and (c), (d) non-crossing estimates of quantile curves for the bone mineral density data: (a), (c) females; (b), (d) males

Takeuchi *et al.* (2006) proposed to minimize the empirical risk plus a regularizer to obtain a non-parametric estimate of the quantile function. As pointed out by Takeuchi *et al.* (2006) the method can be modified to avoid quantile crossing, yielding a quadratic optimization procedure. However, this modification destroys some nice properties of the unconstrained quantile estimate (see the discussion on page 1240 in Takeuchi *et al.* (2006)).

It is the purpose of the present paper to propose an alternative and very simple non-parametric non-crossing estimate of quantile curves. The method starts with an initial non-parametric estimate of the conditional distribution function and solves the problem of determining the solution $\bar{q}_p(x)$ of the equation $p=\hat{F}(y|x)$ and monotonizing the function $p \rightarrow \bar{q}_p(x)$ with respect to the probability p simultaneously. The concept is based on the idea of non-decreasing rearrangements (see Benett and Sharpely (1988)), which was recently used by Dette *et al.* (2006) in the context of estimating a monotone regression. The basic arguments of this technique are carefully described in Section 2. Although the general methodology is applicable to any estimate of the conditional distribution function, we concentrate in Section 3 on the case of independent and identically distributed observations $\{(X_i, Y_i) | i=1, \ldots, n\}$ and on the local constant and local linear estimate of the conditional distribution function (1.3) (see Fan *et al.* (1996)) for definiteness. It is demonstrated that the proposed estimates of the conditional quantile curves are asymptotically normally distributed and share the asymptotic bias and variance of the estimates that were proposed by Yu and Jones (1998) and by Hall *et al.* (1999). In Section 4 some simulation results are presented. We also compare the new procedure with the methods of Yu and Jones (1998) and Hall *et al.* (1999), which are most similar in spirit to the estimate that is proposed in this paper, and with an alternative method that was proposed by He (1997) for location–scale models. Finally, all technical proofs are deferred to Appendix A.

Our results are closely related to the recent and very interesting work of Chernozhukov *et al.* (2007a, b), who independently used the concept of non-decreasing rearrangements to isotonize a preliminary parametric estimate of the conditional quantile function (see also Dette *et al.* (2006) for the application of this concept in the context of isotonic regression). Roughly speaking the procedure that was proposed by Chernozhukov *et al.* (2007a, b) constructs non-crossing estimates of the quantile curves in two steps. First a preliminary (parametric) estimate of the conditional quantile curve is isotonized and inverted. The resulting statistic can be interpreted as an estimate of the conditional distribution function. Secondly, the final non-crossing estimates are constructed by an inversion of the curves that are obtained in the first step. In contrast with Chernozhukov *et al.* (2007a, b), the method that is proposed in this paper consists only of one step and solves the problem of smoothing and inversion of the conditional distribution function simultaneously. We provide the asymptotic theory for non-parametric non-crossing estimates of the quantile distribution and—in contrast with Chernozhukov *et al.* (2007a, b)—our approach does not require a compact support of the conditional distribution function $F(y|x)$. There are several other important differences between both approaches and a detailed discussion will be given in Section 4.

2. Non-crossing non-parametric estimates of quantile curves

Let $\{(X_i, Y_i)\}_{i=1}^n$ denote a bivariate sample of observations with joint density $f_{(X,Y)}$. For simplicity we assume that the explanatory variables are real valued, but the extension to d-dimensional ($d > 1$) covariates will be straightforward. Throughout this paper

$$f_X(u) = \int f_{(X,Y)}(u, v)\,\mathrm{d}v$$

denotes the (marginal) density of the random variables X_i and

$$F_x(y) = F(y|x) = P(Y_i \leqslant y | X_i = x) = \frac{h(x, y)}{f_X(x)} \tag{2.1}$$

is the conditional distribution of Y_i given $X_i = x$. Let $G : \mathbb{R} \rightarrow [0, 1]$ denote a strictly increasing (fixed and known) distribution function. Then the basic idea of our approach can be easily explained as follows: for a distribution function $F : \mathbb{R} \rightarrow \mathbb{R}$ and $p \in F(\mathbb{R})$ we consider the integral

$$H^{-1}(p) := \int_{\mathbb{R}} I\{F(y) \leqslant p\}\,\mathrm{d}G(y) = \int_0^1 I[F\{G^{-1}(v)\} \leqslant p]\,\mathrm{d}v. \tag{2.2}$$

The function G is introduced here to enforce the existence of the integral if the support of the distribution function is unbounded and its choice will be discussed below. The function $p \rightarrow H^{-1}(p)$ is obviously always non-decreasing and for a strictly increasing function F it follows that $H^{-1}(p) = G \circ F^{-1}(p)$. Consequently, in this case we obtain

$$G^{-1} \circ H^{-1}(p) = F^{-1}(p). \tag{2.3}$$

The function H^{-1} is not smooth, but smoothing can easily be accomplished by using

$$H_{h_d}^{-1}(p) = \frac{1}{h_d} \int_{-\infty}^{p} \int_0^1 K_d\left[\frac{F\{G^{-1}(v)\} - u}{h_d}\right] \mathrm{d}v\, \mathrm{d}u \underset{h_d \to 0}{\to} H^{-1}(p), \tag{2.4}$$

where K_d is a symmetric density (with properties that are specified below) and h_d denotes a bandwidth converging to 0 (throughout this paper we assume the existence of the corresponding integrals without further mentioning this fact). The function $H_{h_d}^{-1}$ is always increasing because $(H_{h_d}^{-1})'(p) \geqslant 0, \forall p \in F(\mathbb{R})$, and $G^{-1} \circ H_{h_d}^{-1}$ is an increasing approximation of F^{-1} (provided that this function is invertible).

The construction of an estimate of conditional regression quantiles is now straightforward. For computational reasons the integration with respect to the variable $\mathrm{d}v$ in equation (2.4) is substituted by a summation and the function F is replaced by an appropriate estimate of the conditional distribution, say $\hat{F}_x$, which will be specified below. Under some assumptions of regularity for this estimate the statistic

$$\hat{H}_x^{-1}(p) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{p} K_d\left[\frac{\hat{F}_x\{G^{-1}(i/N)\} - u}{h_d}\right] \mathrm{d}u \tag{2.5}$$

is a consistent estimate of the quantity $G\{F_x^{-1}(p)\}$. This yields

$$\hat{F}_{x,G}^{-1}(p) = (G^{-1} \circ \hat{H}_x^{-1})(p) \tag{2.6}$$

as an estimate of the conditional quantile $F_x^{-1}(p)$, which is (by the reasoning of the previous paragraph) isotonic with respect to p.

Remark 1. For a simple interpretation of this estimate we note that the statistic $\hat{H}_x^{-1}$ is essentially the convolution of the distribution with density $K_d(x/h_d)/h_d$ and the empirical distribution with masses $1/N$ at the points $\{\hat{F}_x\{G^{-1}(i/N)\}\}_{i=1,\ldots,N}$. Therefore the estimate $\hat{H}_x^{-1}$ can be interpreted as a smooth bootstrap estimate of the distribution function of the random variable $\hat{F}_x\{G^{-1}(U)\}$, where the random variable U is uniformly distributed on the interval [0,1].

Remark 2. The estimate (2.6) depends on the given distribution function G. However, for the estimate (1.3) with Nadaraya–Watson or local linear weights it will be demonstrated in the following section that in the case $h_d = o(h_r)$ this dependence is not visible in the first-order asymptotics. Similar observations can also be made for other estimates of the conditional distribution function such as local polynomials or smoothing spline methods (see Masry and Fan (1997) and Gu (1995)). Also it is reasonable to choose the function G in dependence on the explanatory variable x. Some recommendations regarding the choice of G will be given in Section 4 and it will be demonstrated by means of a simulation study that even for finite sample sizes the effect of the function G on the statistical properties of the resulting estimate $\hat{F}_{x,G}$ is nearly non-visible.

Remark 3. Definition (2.6) does not require a local linear or Nadaraya–Watson type of estimate of the conditional distribution function. In fact the procedure can be used to obtain non-crossing quantile curves from any parametric or non-parametric estimate of the conditional distribution function. However, the theoretical (asymptotic) properties of the estimate obviously depend on the initial estimate of $F(y|x)$. In the following section we derive such properties for the estimate (1.3) with Nadaraya–Watson and local linear weights and indicate the corresponding results if alternative estimates are used for the conditional distribution function. For more details see remark 7 in Section 3.

3. Asymptotic results

For definiteness we consider a sample of independent and identically distributed observations $(X_1, Y_1), (X_2, Y_2), \ldots$, but the results can easily be extended to sequences of dependent random variables. As an estimate of the conditional distribution function we use the statistic $\hat{F}_x(y) = \hat{F}(y|x)$ that is defined in equation (1.3), where the weights $w_i(x)$ are either the Nadaraya–Watson

$$w_i(x) = K_r\left(\frac{x - X_i}{h_r}\right) \Big/ \sum_{j=1}^{n} K_r\left(\frac{x - X_i}{h_r}\right), \qquad i = 1, \ldots, n, \tag{3.1}$$

or the local linear weights

$$w_i(x) = K_r\left(\frac{x - X_i}{h_r}\right) \{S_{n,2} - (x - X_i) S_{n,1}\} \Big/ (S_{n,2} S_{n,0} - S_{n,1}^2), \qquad i = 1, \ldots, n, \tag{3.2}$$

with

$$S_{n,l} = \sum_{j=1}^{n} K_r\left(\frac{x - X_i}{h_r}\right) (x - X_j)^l, \qquad l = 0, 1, 2, \tag{3.3}$$

(see Fan and Gijbels (1996)). In equations (3.1) and (3.2) the quantity K_r denotes a symmetric, continuous kernel with compact support, say $[-1, 1]$, and h_r refers to the corresponding bandwidth tending to 0 with increasing sample size. The kernel K_d (see equation (2.4) for its role) in the statistic is assumed to be twice continuously differentiable, symmetric and positive with compact support $[-1, 1]$, and the corresponding bandwidth h_d converges to 0 with increasing sample size. Furthermore, we assume for the bandwidths h_d and h_r

$$\left.\begin{array}{r} n h_r^5 = c + o(1), \\ h_d = o(h_r), \\ h_r^4 / h_d^3 = o(1) \end{array}\right\} \tag{3.4}$$

for some positive constant c, and that the sample size satisfies $N = O(n)$. Our main results specify the asymptotic distribution of the corresponding quantile estimate that is obtained from equation (2.6) with equation (1.3) as initial estimate. The proof is deferred to Appendix A. Throughout this paper $\partial_1^k s$ and $\partial_2^k s$ denote the kth partial derivative of the function $s(x, y)$ or $s(y|x)$ with respect to the co-ordinate x and y respectively, and for a kernel K we define $\mu_2(K) = \frac{1}{2} \int_{-1}^{1} t^2 \, K(t) \, \mathrm{d}t$.

Theorem 1. Assume that the functions $f_{(X,Y)}$, F_x and G are twice continuously differentiable and that the function

$$h(x, y) = \int_{-\infty}^{y} f_{(X,Y)}(x, v) \, \mathrm{d}v \tag{3.5}$$

is twice differentiable with respect to the first co-ordinate such that for any x the function $\partial_1^2 h(x, y)$ is Lipschitz continuous with respect to y and satisfies

$$\sup_{z \in U_\varepsilon(x), t \in \mathbb{R}} |\partial_1^2 h(z, t)| < \infty \tag{3.6}$$

in some neighbourhood $U_\varepsilon(x)$ of x. Suppose further that

$$\int |f_X''(u)|\,\mathrm{d}u < \infty,$$
$$\int\int |\partial_1 f_{(X,Y)}(u,v)|\,\mathrm{d}u\,\mathrm{d}v < \infty \tag{3.7}$$

and let $p \in (0,1)$ be such that $F_x'\{F_x^{-1}(p)\} > 0$ and $G'\{F_x^{-1}(p)\} > 0$.

(a) If the weights in equation (1.3) are given by the Nadaraya–Watson weights that are defined in equation (3.1), then for any $p \in (0,1)$ the corresponding quantile estimate (2.6) converges weakly, i.e.

$$\surd(nh_r)\{\hat{F}_{x,G}^{-1}(p) - F_x^{-1}(p) + b_n(x,p)\} \xrightarrow{\mathcal{D}} \mathcal{N}\{0, V(x,p)\},$$

with bias

$$b_n(x,p) = \frac{h_r^2\,\mu_2(K_r)}{f_X(x)\,F_x'\{F_x^{-1}(p)\}}[\partial_1^2 h\{x, F_x^{-1}(p)\} - f_X''(x)p], \tag{3.8}$$

and asymptotic variance

$$V(x,p) = \frac{p(1-p)\int K_r^2(u)\,\mathrm{d}u}{f_X(x)\,F_x'\{F_x^{-1}(p)\}^2}. \tag{3.9}$$

(b) If the weights in equation (1.3) are given by the local linear weights that are defined in equation (3.2), then for any $p \in (0,1)$ the corresponding quantile estimate (2.6) converges weakly, i.e.

$$\surd(nh_r)\{\hat{F}_{x,G}^{-1}(p) - F_x^{-1}(p) + \bar{b}_n(x,p)\} \xrightarrow{\mathcal{D}} \mathcal{N}\{0, V(x,p)\}$$

with bias

$$\bar{b}_n(x,p) = \frac{h_r^2\,\mu_2(K_r)}{F_x'\{F_x^{-1}(p)\}}\partial_1^2 F\{F_x^{-1}(p)|x\}$$

and asymptotic variance $V(x,p)$ defined by equation (3.9).

Remark 5. The estimate that is considered in the second part of theorem 1 has the same asymptotic distribution as the local double-kernel quantile estimate that was proposed by Yu and Jones (1998) if the corresponding second bandwidth in this estimate is of smaller order than the first bandwidth (see theorem 1 in Yu and Jones (1998)). However, for the local constant and local linear quantile estimates that are defined by equations (1.1) and (1.2), it follows from the results in Yu and Jones (1997) that the bias and variance of the local constant fit (1.1) are given by

$$h_r^2\,\mu_2(K_r)\left[-\frac{\partial_1^2 F\{F_x^{-1}(p)|x\}}{F_x'\{F_x^{-1}(p)\}} + 2\frac{f_X'(x)\,F_x^{-1}(p)'}{f_X(x)}\right] + o(h_r^2),$$
$$\frac{p(1-p)}{nh_r\,F_x'\{F_x^{-1}(p)\}^2}\int K_r^2(u)\,\mathrm{d}u + o\left(\frac{1}{nh_r}\right)$$

respectively. Similarly, the local linear estimate that is defined by equation (1.2) yields a bias of the form

$$h_r^2\,\mu_2(K_r)\,F_x^{-1}(p)'' + o(h_r^2)$$

and the same variance. Thus, there is a difference between the estimates that are proposed in this paper and the methods (1.1) and (1.2) that are based on a combination of the concept of a check function and local smoothing. However, the important difference for practical applications is that the new estimates keep the logical monotonicity restrictions with respect to the probability p without any additional modification.

Remark 6. It is worthwhile to mention that the inverse of the statistic $\hat{F}^{-1}_{x,G}$, say

$$\hat{F}_{x,G}(y) = \hat{H}_x\{G(y)\},$$

yields a strictly increasing estimate of the conditional distribution function $F(y|x)$, which satisfies the constraints $0 \leqslant \hat{F}_x(y) \leqslant 1$ $(y \in \mathbb{R})$. In this sense the methodology proposed can also be viewed as an alternative to the adjusted Nadaraya–Watson estimate that was suggested by Hall *et al.* (1999). Moreover, if $h_d = o(h_r)$, it follows for the estimate (1.3) with local linear weights by similar arguments to those presented in the proof of theorem 3.2 in Dette *et al.* (2006) that

$$\surd(nh_r)\{\hat{F}_{x,G}(y) - F(y|x) - c_n(y|x)\} \xrightarrow{\mathcal{D}} \mathcal{N}\{0, w(y|x)\},$$

where

$$c_n(y|x) = \frac{h_r^2 \mu_2(K_r)}{f_X(x)} \partial_1^2 F(y|x)$$

and

$$w(y|x) = \frac{F(y|x)\{1 - F(y|x)\}}{f_X(x)} \int K^2(u)\, \mathrm{d}u.$$

This shows that the estimate $\hat{F}_{x,G}$ has the same asymptotic distribution as the adjusted Nadaraya–Watson estimate that was proposed by Hall *et al.* (1999). It is also worthwhile to mention that in contrast with the method that was suggested by Hall *et al.* (1999) the estimate that is proposed in this paper can be easily extended to local polynomial techniques (see Fan and Gijbels (1996) or Masry and Fan (1997)) without loosing monotonicity and the properties of a distribution function.

Remark 7. As indicated in Section 1 the procedure that was proposed in Section 2 is applicable to any parametric or non-parametric estimate of the conditional distribution function and similar results to those presented in theorem 1 can be obtained provided that there is a limit theorem for the initial estimate of the conditional distribution function, where the bias and variance in theorem 1 must be modified appropriately. In fact one can (under some technical assumptions) show that the weak convergence of

$$a_n[\hat{F}_x\{F_x^{-1}(p)\} - F\{F_x^{-1}(p)|x\} - b_n\{F_x^{-1}(p)|x\}] \xrightarrow{\mathcal{D}} \mathcal{N}[0, \sigma^2\{F_x^{-1}(p)|x\}]$$

implies that the corresponding quantile estimator that is defined in Section 2, say $\hat{F}^{\text{inv}}_{x,G}(p)$, converges in law, i.e.

$$a_n\left[\hat{F}^{\text{inv}}_{x,G}(p) - F_x^{-1}(p) + \frac{b_n\{F^{\text{inv}}_{x,G}(p)|x\}}{F_x'\{F_x^{-1}(p)\}}\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left[0, \frac{\sigma^2\{F_x^{-1}(p)|x\}}{F_x'\{F_x^{-1}(p)\}^2}\right]$$

(see Volgushev (2007)). An important non-parametric example is the local polynomial estimator of odd degree r (see Fan and Gijbels (1996) for details). If the corresponding weights are used for the estimation of the conditional distribution function we obtain

$$\surd(nh_r)\{\hat{F}^{\text{inv}}_{x,G}(p) - F_x^{-1}(p) + \tilde{b}_n(p|x)\} \xrightarrow{\mathcal{D}} \mathcal{N}\{0, \tilde{\sigma}^2(p|x)\}$$

with asymptotic bias and variance

$$\tilde{b}_n(p|x) = e_1^t S^{-1} c_r \frac{1}{(r+1)!} \frac{\partial_1^r F\{F_x^{-1}(p)|x\}}{F_x'\{F_x^{-1}(p)\}} h^{r+1},$$

$$\tilde{\sigma}^2(p|x) = e_1^t S^{-1} S^* S^{-1} e_1 \frac{p(1-p)}{f_X(x)\ F_x'\{F_x^{-1}(p)\}^2},$$

where we use the notation

$$S := \left(\int u^{j+l}\ K_r(u)\ \mathrm{d}u\right)_{0 \leqslant j,l \leqslant r},$$

$$S^* := \left(\int u^{j+l}\ K_r^2(u)\ \mathrm{d}u\right)_{0 \leqslant j,l \leqslant r},$$

$c_r := (\int u^{r+l}\ K_r(u)\ \mathrm{d}u, \ldots, \int u^{2r+l}\ K_r(u)\mathrm{d}u)$ and $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^{r+1}$. Thus the method that is proposed in this paper yields non-crossing estimates of quantile curves with a higher order expansion, similar to the usual local polynomial estimate in non-parametric regression.

Remark 8. As indicated in remark 2 a data-adaptive choice of the function G, say G_n, can improve the finite sample properties of the estimate. Note that the asymptotic properties of the estimate of the conditional quantiles do not depend on the (fixed) choice of the function G in equation (2.6) which satisfies the assumptions of theorem 1. Therefore, if the function G in equation (2.6) is replaced by a data-adaptive estimate, which converges to G sufficiently fast, it can be shown that theorem 1 is still valid. For example, in the finite sample analysis of the proposed estimate presented in Section 5 we fix a function G, consider the location–scale family $G\{(x-a)/b\}$ and determine the location and scale estimate a_n and b_n from the data, such that the 0.05- and 0.95-quantiles of the resulting distribution function $G_n(x) = G\{(x-a_n)/b_n\}$ coincide with the corresponding empirical quantile of the sample $\{Y_j||X_j - x| < h_d;\ j = 1, \ldots, n\}$. It follows that a_n and b_n are consistent estimates of the parameters a_0 and b_0, which are defined by the equations

$$G^{-1}(0.05) = \frac{F_x^{-1}(0.05) - a_0}{b_0},$$

$$G^{-1}(0.95) = \frac{F_x^{-1}(0.95) - a_0}{b_0},$$

and as a consequence the sequence $G_n(\cdot)$ converges to $G\{(\cdot - a_0)/b_0\}$. Now going through the proofs in Appendix A and replacing G by G_n shows that theorem 1 is still valid if the distribution function G is replaced by its data-dependent version G_n. The details are omitted for brevity.

Remark 9. The main practical applications of the procedure proposed appear for univariate predictors, because in the case of a multivariate predictor the initial non-parametric estimate of the conditional distribution suffers from the curse of dimensionality. In this case the consideration of parametric models seems to be more realistic from a practical point of view. It is worthwhile to mention that the proposed procedure of simultaneous inversion and monotonization can also be applied to an initial parametric estimate of the conditional distribution function (see remark 7).

4. Non-decreasing rearrangements of quantile estimates

Our results are closely related to the recent and very interesting work of Chernozhukov *et al.*

(2007a, b), who provided asymptotic properties of non-crossing parametric estimates of conditional quantiles. They applied a (non-smooth) version of non-decreasing rearrangements that were proposed by Dette *et al.* (2006) for isotonic regression to obtain non-crossing quantile estimates from initial parametric quantile estimates. Roughly speaking, for a conditional distribution function $F(y|x)$ they considered the (non-smoothed) estimate

$$\tilde{F}(y|x)=\int_0^1 I\{\hat{Q}(u|x)\leqslant y\}\,\mathrm{d}u, \tag{4.1}$$

where $\hat{Q}(u|x)$ is a preliminary parametric estimate of the conditional u-quantile, which does not necessarily yield non-crossing estimates of quantile curves. Because $\tilde{F}(y|x)$ is obviously monotone with respect to y, non-crossing quantile curves are easily obtained from the inverse of $\tilde{F}(y|x)$, i.e.

$$\tilde{F}^{-1}(p|x)=\inf\{y|\tilde{F}(y|x)\geqslant p\}. \tag{4.2}$$

In other words, the estimate of Chernozhukov *et al.* (2007a, b) is a two-step procedure. In the first step a monotone estimate of the conditional distribution function is obtained from the initial quantile estimate $\hat{Q}$ and in the second step the inverse of this function is calculated.

Chernozhukov *et al.* (2007a) showed that the rearranged curve $\tilde{F}^{-1}(p|x)$ is always closer to the true curve than the preliminary estimate $\hat{Q}(p|x)$, where distances are measured with respect to the L^q-norm for $q\geqslant 1$. This extends recent findings of Birke and Dette (2007), who showed that $\tilde{F}^{-1}(p|x)$ and $\hat{Q}(p|x)$ have the same L^q-norm for all $q\geqslant 1$. Moreover, Chernozhukov *et al.* (2007b) proved the Hadamard differentiability of the operator that is defined in equation (4.1) and this result is finally applied to obtain the asymptotic distribution for parametric non-crossing estimates of the conditional quantiles. A more detailed discussion of the main differences between the results that are presented in this paper and the work of Chernozhukov *et al.* (2007a, b) is given in the following paragraphs.

(a) The operator $H_{h_d}^{-1}$ that is defined in equation (2.5) converges to the operator

$$\tilde{H}^{-1}(p)=\int_0^1 I\{\hat{F}_x(u)\leqslant p\}\,\mathrm{d}u \tag{4.3}$$

if $h_d\to 0$ and $N\to\infty$, and G corresponds to the uniform distribution on the interval [0,1]. Chernozhukov *et al.* (2007b) used this operator to invert and isotonize preliminary estimates of the quantile curves. The resulting statistic is an estimate of the conditional distribution function and its inverse is calculated in the second step. In contrast with Chernozhukov *et al.* (2007a, b) we apply a (smooth) version of operator (4.3) directly to a non-parametric estimate of the conditional distribution function to obtain non-crossing non-parametric quantile estimates. Therefore the procedure that is proposed in this paper consists only of one step and solves the problem of inversion and isotonization of the estimate of the conditional distribution function simultaneously.

(b) The main results in the work of Chernozhukov *et al.* (2007b) require a compact support of the conditional distribution function $F(y|x)$. By introducing the additional distribution function G we can establish a methodology which is applicable to conditional distribution functions with an unbounded support. This point is of particular importance if extreme quantiles are of interest to the experimenter.

(c) In constast with Chernozhukov *et al.* (2007b) we provide asymptotic results for the weak convergence of a non-parametric (non-crossing) estimate under the assumption of a strictly increasing distribution, i.e. $\partial F(p|x)/\partial p>0$. Chernozhukov *et al.* (2007b) concentrated on the parametric model and discussed some results in the case where the set $\{p \in (0,1) | \partial F(p|x)/\partial p = 0\}$ is finite. Of course this cannot occur when the parametric model is specified correctly; however, it has an interesting implication in the case of model misspecification (for details see Chernozhukov *et al.* (2007b)).

(d) It was demonstrated by Dette *et al.* (2006) and Chernozhukov *et al.* (2007b) that the resulting quantile estimates are not necessarily smooth with respect to p (see Fig. 1 in Chernozhukov *et al.* (2007b)). The method that is proposed in this paper provides smooth and non-crossing estimates of the conditional quantile function. Moreover, it was pointed out by Dette *et al.* (2006) that the choice of the smoothing parameter h_d is not very critical as long as it is chosen sufficiently small (see also our numerical study in Section 4).

(e) Chernozhukov *et al.* (2007b) considered arbitrary preliminary (parametric) estimates of the conditional quantile curve. Similarly, the results of the present paper are valid for any preliminary non-parametric estimate of the conditional distribution function, provided that it has similar properties to those of the Nadaraya–Watson or local linear estimate that was considered in Section 3. In fact our results may also be applied to initial parametric estimates of the conditional distribution function, provided that these estimates satisfy some technical assumptions (see remark 7).

5. Empirical results

In the present section we study the finite sample properties of the new estimate by means of a simulation study. First we investigate the effect of the choice of the function G on the quantile estimate. Secondly we compare the method that is proposed in this paper with some of the currently available procedures for non-crossing quantile curves (see He (1997), Yu and Jones (1998) and Hall *et al.* (1999)).

5.1. Sensitivity with respect to choice of G

Recall from the asymptotic results that were presented in Section 3 that for the choice $h_d = o(h_r)$ the function G has no effect on the asymptotic properties of the new estimates for the quantile curves. For practical applications it is of interest to investigate in what sense these observations can be transferred to realistic sample sizes.

For finite sample sizes the function G should depend on the point x where we want to estimate the quantile $F^{-1}(p|x)$. This is intuitively clear, but it also follows from the definition (2.5) in Section 2 that the distribution function of G should give a reasonable weight to most of the responses $Y_1, \ldots, Y_n$ whose corresponding predictors are close to the point x. Thus in practice the function G should be adapted to the point x and the data, and we propose the following method for this purpose. In the first step we define for fixed x the set

$$\{Y_j || X_j - x| < h_d; j = 1, \ldots, n\} \tag{5.1}$$

of all observations Y_j with a corresponding predictor in the neighbourhood of x, and we denote its ordered elements by $Z_1, \ldots, Z_k$. We consider a distribution from a location–scale family and choose $G = \hat{G}_x$ such that $\hat{G}_x^{-1}(0.05) = Z_{\lfloor 1+0.05k \rfloor}$ and $\hat{G}_x^{-1}(0.95) = Z_{\lfloor 0.95k \rfloor}$. To study the effect of the form of G on the resulting estimate we investigate three distributions: the Cauchy, double-exponential and normal distributions, where the scaling and location parameters have been adapted to the data by the procedure that was described in the previous paragraph. In Fig. 2 we show the three estimates (which were obtained from the different choices for the function G) for examples 2 and 3 in Yu and Jones (1998), i.e.

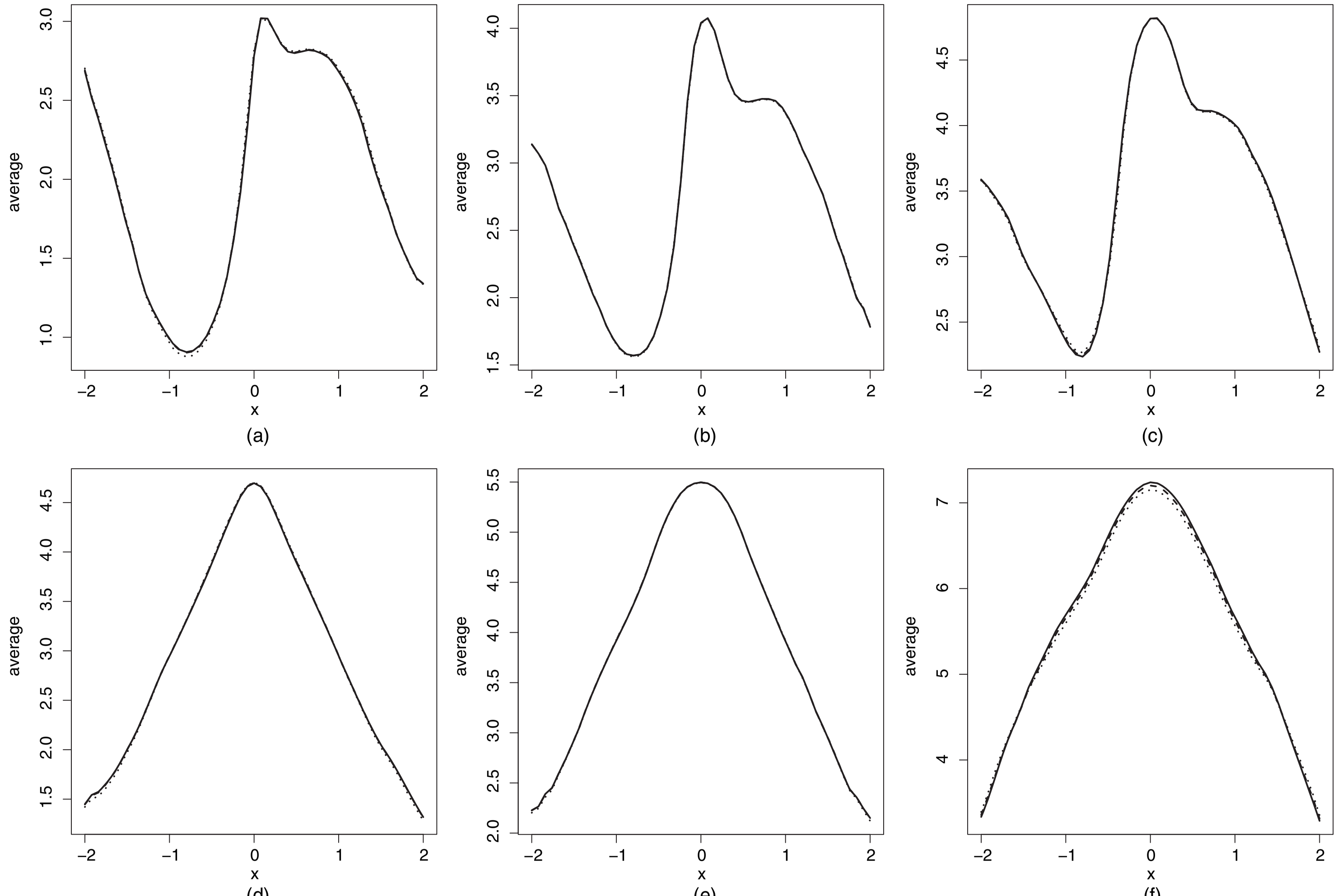

Fig. 2. Non-parametric estimates of the quantile curves functions G and various quantiles for (a)–(c) model (5.2) and (d)–(f) model (5.3) (the estimates result from the use of the normal (———), double-exponential (- - - - -) and Cauchy distributions (· · · · · · ·)): (a), (d) $p=0.1$; (b), (e) $p=0.5$; (c), (f) $p=0.9$

$$Y = 2.5 + \sin(2X) + 2\exp(-16X^2) + 0.5Z, \qquad X \sim \mathcal{N}(0,1), \quad Z \sim \mathcal{N}(0,1), \tag{5.2}$$

$$Y = 2 + 2\cos(X) + \exp(-4X^2) + E, \qquad X \sim \mathcal{N}(0,1), \quad E \sim \exp(1). \tag{5.3}$$

Fig. 2 shows the resulting estimates of the 10%, 50% and 90% conditional quantile curves averaged over 100 replications. For the bandwidth h_r in the estimate of the conditional distribution function we apply the bandwidth selection rule that was proposed in Yu and Jones (1998),

$$h_r = h_r(p) = \hat{h}_{r,\text{opt}} \left[\frac{p(1-p)}{\varphi\{\Phi^{-1}(p)\}^2} \right]^{1/5}, \tag{5.4}$$

where φ and Φ denote the density and distribution function of the standard normal distribution respectively, and we use the method of Ruppert *et al.* (1995) to obtain the preliminary estimate $\hat{h}_{r,\text{opt}}$. All simulations were done in the statistical software R (R Development Core Team, 2006) and we used the package 'Kern Smooth' (Wand and Ripley, 2005) to compute the bandwidth. The kernel K_r was chosen as a Gauss kernel, whereas the Epanechnikov kernel was used for K_d. The sample size was $n = 100$ and we used $N = 1000$ points to evaluate the function $\hat{G}_x^{-1}$ in equation (2.5). Too large or too small values of the estimate $\hat{F}_{x,G}^{-1}(p)$ would yield a numerically unstable procedure which could destroy the basic properties of the estimate. For this reason the statistic $\hat{H}_x^{-1}(p)$ was truncated to lie in the interval [0.05,0.95]; more precisely, we used $0.95\, I\{\hat{H}_x^{-1}(p) > 0.95\} + \hat{H}_x^{-1}(p)\, I\{0.05 < \hat{H}_x^{-1}(p) < 0.95\} + 0.05\, I\{\hat{H}_x^{-1}(p) < 0.05\}$ instead of $\hat{H}_x^{-1}(p)$. The choice of the bandwidth h_d is less critical and we used

$$h_d = \min(p, 1-p, h_r^{1.3}). \tag{5.5}$$

The results are displayed in Fig. 2 and in both models we observe no visible differences between the three estimates that were obtained by different choices for the function G. Further simulation results investigating the effect of the choice of the function G are available from the second author and demonstrate that the results that are presented in Fig. 2 are representative for a large number of situations. If the function G is adapted to the data and the point x according to the procedure that was described at the beginning of this section or according to a similar method, the particular form of the function G has no visible effect on the estimate of the quantile curve for realistic sample sizes.

5.2. Finite sample comparison

We now compare the new estimate with the procedures that were proposed by Yu and Jones (1998) and Hall *et al.* (1999), which are most similar in spirit to the method that is suggested in this paper. For brevity we restrict ourselves to models (5.2) and (5.3). Yu and Jones (1998) considered two other models for which similar results are available from the authors. In Fig. 3 we display the square root of the mean-squared error of the estimates of Yu and Jones (1998) (dotted curve), Hall *et al.* (1999) (broken curve) and the method that is proposed in this paper (full curve). All estimates require a bandwidth for the regression step, which is chosen by equation (5.4). The method of Yu and Jones (1998) requires a further bandwidth, which is chosen according to their formula (12), whereas the bandwidth h_d in the new procedure is given by equation (5.5) and the function G by the distribution function of a normal distribution. We display the (pointwise) square root of the mean-squared error that is obtained from 1000 simulation runs for the quantiles $p = 10\%$, $p = 50\%$ and $p = 90\%$. In model (5.2) all three estimates have similar behaviour, where the new method yields a slightly smaller mean-squared error (see

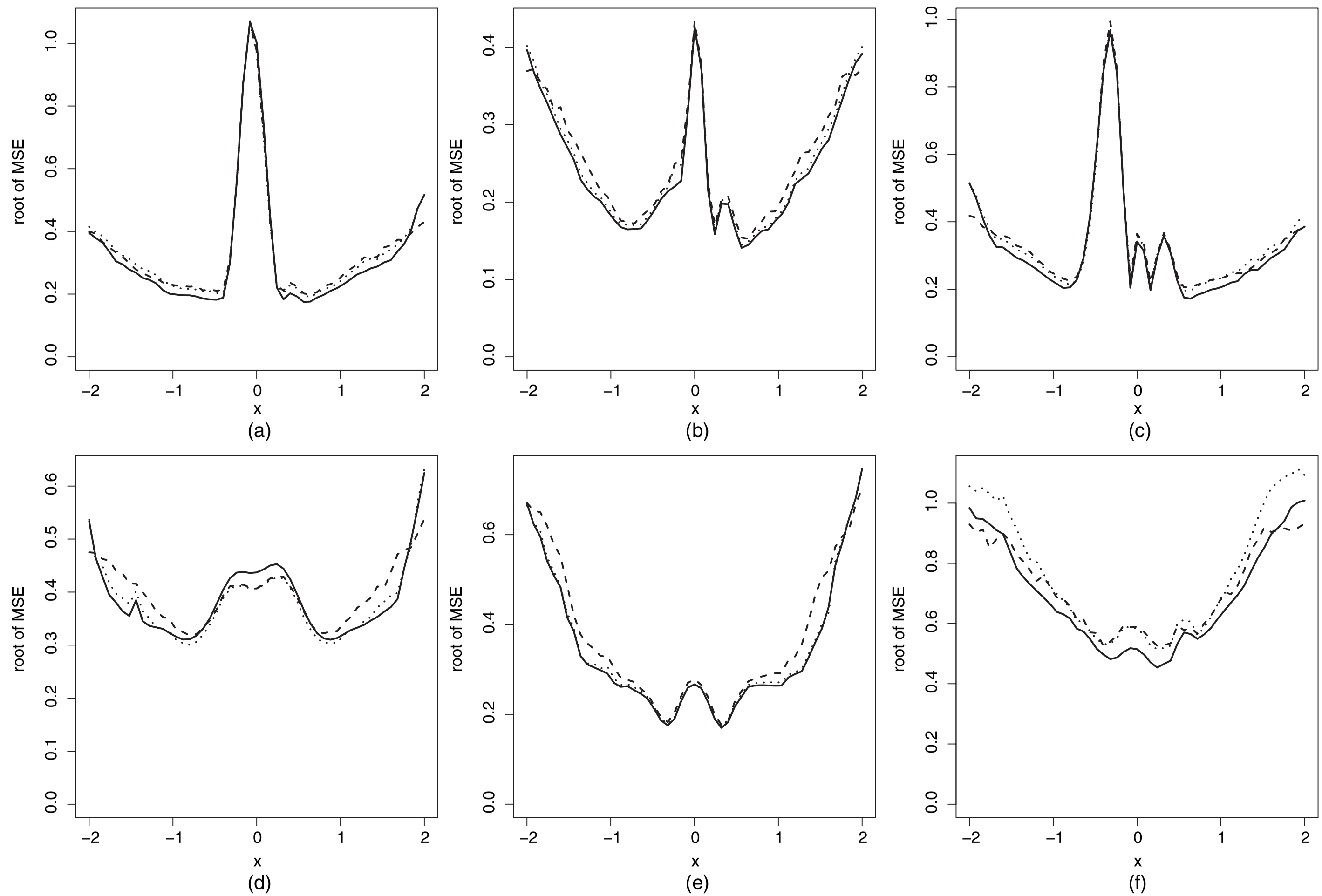

Fig. 3. Square root of the mean-squared error of non-parametric estimates in (a)–(c) model (5.2) and (d)–(f) model (5.3) for various quantile estimates (– – –, estimate of Hall *et al.* (1999); · · · · · · ·, estimate of Yu and Jones (1998); ———, estimate $\hat{F}_{X,G}$ proposed in this paper): (a), (d) $p=0.1$; (b), (e) $p=0.5$; (c), (f) $p=0.9$

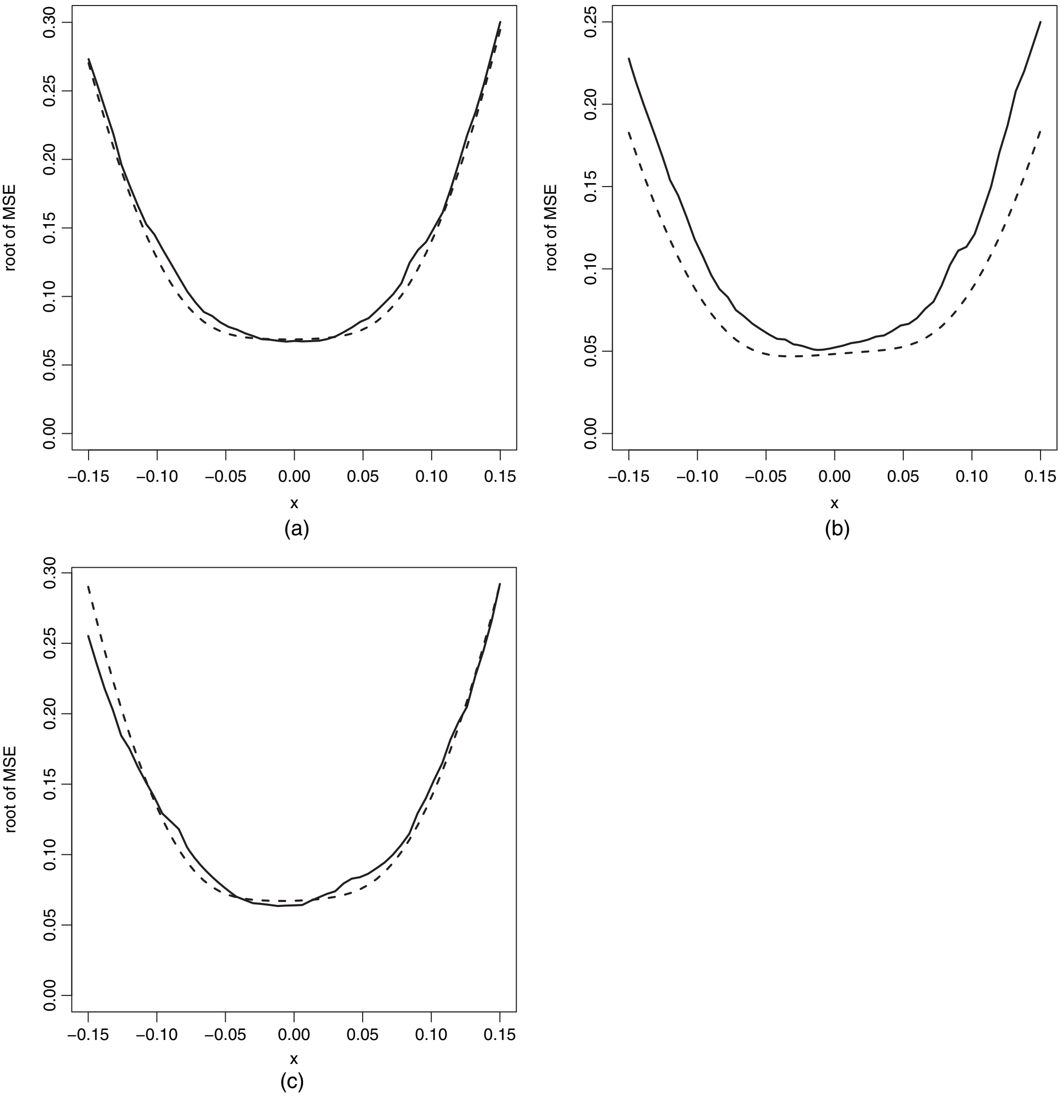

Fig. 4. Square root of the mean-squared error of the estimate of He (1997) (- - - - -) and the estimates from this paper (———) in a location–scale model: (a) $p = 0.1$; (b) $p = 0.5$; (c) $p = 0.9$

Figs 3(a)–3(c)). In model (5.3) the estimate of Hall *et al.* (1999) yields a larger mean-squared error when estimating the 10% and 50% conditional quantiles whereas the estimate of Yu and Jones (1998) and the new method show a very similar performance. The situation is quite different when estimating the 90% quantile curve; here the estimate of Yu and Jones (1998) yields the largest mean-squared error whereas the estimator of Hall *et al.* (1999) has some advantage over the new method and the estimate of Yu and Jones (1998) near the boundary. In the interior part of the design region the new method always yields a smaller mean-squared error. Summarizing these results (and further simulations, which for brevity are not displayed) we note that the estimate of Hall *et al.* (1999) usually yields a larger mean-squared error than the estimates of Yu and Jones (1998) and the estimates that are proposed in this paper. These two estimates show very similar behaviour with some advantages for the new method.

It was pointed out by a referee that it is of interest to compare the new estimate with an alternative method which has been proposed by He (1997) for constructing non-crossing quantile curves in location–scale models, i.e.

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \qquad i = 1, \ldots, n. \tag{5.6}$$

He (1997) proposed to estimate the conditional median function $\hat{m}$ of Y given X and to compute the residuals $R_i = Y_i - \hat{m}(X_i)$. In the next step a conditional median regression is applied to the pairs $(\{X_i, |R_i|\})$ to estimate the conditional variance $\hat{\sigma}(x)$ (this estimate must be constrained to be positive) and the normalized residuals $V_i = R_i/\hat{\sigma}(X_i)$ are calculated. Because of the location–scale structure, these residuals should approximately have the same distribution as the errors ε_i. Hence, from the empirical α-quantile c_α of the sample $V_1, \ldots, V_n$, an estimate of the conditional α-quantile $q_\alpha(x)$ of Y_i given $X = x$ can be obtained by $\hat{m}(x) + \hat{\sigma}(x)c_\alpha$. To carry out the median regressions we use quadratic B-splines as suggested by He (1997), and the splines are computed by using the function `cobs` from the R-package 'cobs' by Ng and Maechler (2006) with automatic knot selection.

For illustration we display in Fig. 4 the square root of the mean-squared error in the model

$$Y = 1 + \sin\left(\frac{3}{4}X\right) + 0.3Z\sqrt{\left\{\sin\left(\frac{3}{4}X\right) + 1\right\}}, \qquad X \sim \mathcal{N}\left(0, \frac{1}{16}\right), \quad Z \sim \mathcal{N}(0,1).$$

We observe some advantages of the method of He (1997), which explicitly uses the structure of the location–scale model. However, if assumption (5.6) does not hold, the estimate of He (1997) yields a substantial estimation error. For brevity these results are not displayed but are available from the authors.

Acknowledgements

The authors are grateful to Isolde Gottschlich, who typed numerous versions of this paper with considerable technical expertise, and R. Koenker, who pointed out that the problem of crossing estimates of conditional quantile functions is of importance. The authors also thank two referees and the Associate Editor for their constructive comments on an earlier version of this paper and for drawing our attention to Chernozhukov *et al.* (2007a, b). We further thank V. Chernozhukov for helpful discussions. The work of the authors was supported by the Sonderforschungsbereich 475, 'Komplexitätsreduktion in multivariaten Datenstrukturen' (Teilprojekt B1). The work of H. Dette was supported in part by National Institutes of Health grant award IR01GM072876:01A1.

Appendix A: Proofs

For brevity we restrict ourselves to a proof of the first part of theorem 1; all other results are proved similarly and the arguments are therefore omitted. Moreover, for simple notation we assume that $n = N$. The proof is performed in two steps. In the first step we establish the weak convergence

$$\sqrt{(nh_r)}\{\hat{H}_x^{-1}(p) - H_x^{-1}(p) + \tilde{b}_n(x,p)\} \xrightarrow{\mathcal{D}} \mathcal{N}\{0, \tilde{V}(x,p)\}, \tag{A.1}$$

where

$$\tilde{b}_n(x,p) = \frac{h_r^2\,\mu_2(K_r)}{f_X(x)}[\partial_1^2 h\{x, F_x^{-1}(p)\} - f_X''(x)p]\frac{G'\{F_x^{-1}(p)\}}{F_x'\{F_x^{-1}(p)\}}, \tag{A.2}$$

$$\tilde{V}(x,p) = \frac{p(1-p)\int K^2(u)\,\mathrm{d}u\; G'\{F_x^{-1}(p)\}^2}{f_X(x)\;F_x'\{F_x^{-1}(p)\}^2}, \tag{A.3}$$

and we use the notation $H_x = F_x \circ G^{-1}$. In the second step we apply a Taylor series expansion to complete the proof of theorem 1. More precisely we obtain for the statistic

$$\begin{aligned}\sqrt{(nh_r)}T_n &= \sqrt{(nh_r)}\{\hat{F}_{x,G}^{-1}(p) - F_x^{-1}(p) + b_n(x, p)\} \\ &= \frac{\sqrt{(nh_r)}}{G'\{F_x^{-1}(p)\}}\sqrt{(nh_r)}\{\hat{H}_x^{-1}(p) - H_x^{-1}(p) + \tilde{b}_n(x, p)\} + R_n,\end{aligned}$$

where the remainder term is defined by

$$R_n = \sqrt{(nh_r)}\{\hat{H}_x^{-1}(p) - H_x^{-1}(p)\}\{(G^{-1})'(\xi) - (G^{-1})' H_x^{-1}(p)\}$$

for some ξ satisfying

$$|\xi - H_x^{-1}(p)| \leqslant |\hat{H}_x^{-1}(p) - H_x^{-1}(p)| \xrightarrow{\mathrm{P}} 0.$$

A standard argument and expression (A.1) now show that $R_n = o_p(1)$, and the assertion of the theorem follows if expression (A.1) can be established.

For a proof of the weak convergence in expression (A.1) we introduce the decomposition

$$\hat{H}_x^{-1}(p) - H_x^{-1}(p) = \Delta_n^{(1)}(p) + \tfrac{1}{2}\Delta_n^{(2)}(p) + \Delta_n^{(3)}(p), \tag{A.4}$$

where

$$\Delta_n^{(1)}(p) = -\frac{1}{nh_d}\sum_{i=1}^{n} K_d\left\{\frac{F_x(g_i) - p}{h_d}\right\}\{\hat{F}_x(g_i) - F_x(g_i)\}, \tag{A.5}$$

$$\Delta_n^{(2)}(p) = -\frac{1}{nh_d^2}\sum_{i=1}^{n} K_d'\left(\frac{\xi_i - p}{h_d}\right)\{\hat{F}_x(g_i) - F_x(g_i)\}^2, \tag{A.6}$$

$$\Delta_n^{(3)}(p) = \frac{1}{nh_d}\int_{-\infty}^{t}\sum_{i=1}^{n} K_d\left\{\frac{H_x(i/n) - u}{h_d}\right\}\mathrm{d}u - H_x^{-1}(p), \tag{A.7}$$

and we have used the notation $g_i := G^{-1}(i/n)$ and $|\xi_i - F_x(g_i)| \leqslant |F_x(g_i) - \hat{F}_x(g_i)|$ $(i = 1, \ldots, n)$. From lemma 2.1 in Dette *et al.* (2006) it follows that

$$\Delta_n^{(3)}(y) = \mu_2(K_d)\, h_d^2 (H_x^{-1})''(t) + O\left(\frac{1}{nh_d}\right), \tag{A.8}$$

and assertion (A.1) is now a consequence of Slutzky's lemma if we establish that

$$\sqrt{(nh_r)}\{\Delta_n^{(1)}(p) + \tilde{b}_n(x, p)\} \xrightarrow{\mathcal{D}} \mathcal{N}\{0, \tilde{V}(x, p)\}, \tag{A.9}$$

$$\sqrt{(nh_r)}\;\Delta_n^{(2)}(p) = o_p(1). \tag{A.10}$$

For this purpose we recall the definition of estimate (1.3) for the Nadaraya–Watson weights (3.1) and obtain

$$\Delta_n^{(1)}(p) = -\frac{1}{nh_d}\sum_{i=1}^{n} K_d\left\{\frac{F_x(g_i) - p}{h_d}\right\} D_n(p, g_i)\{1 + o_p(1)\}, \tag{A.11}$$

$$\Delta_n^{(2)}(p) = -\frac{1}{nh_d^2}\sum_{i=1}^{n} K_d'\left(\frac{\xi_i - p}{h_d}\right) D_n^2(x, g_i)\{1 + o_p(1)\}, \tag{A.12}$$

where the quantity $D_n(x, y)$ is a sum of independent and identically distributed random variables, i.e.

$$D_n(x, y) = \sum_{i=1}^{n} Z_{n,i}(y)$$

with

$$Z_{n,i}(y) = \frac{1}{nh_r}\frac{1}{f_X(x)}K_r\left(\frac{X_i - x}{h_r}\right)\{I_{\{Y_i \leqslant y\}} - F_x(y)\}.$$

Consequently, the leading term in equation (A.11) can be written as $H_x^{\text{lin}}(p) = \Sigma_{k=1}^n Y_{n,k}(p)$, where the independent, identically distributed random variables $Y_{n,k}(p)$ are given by

$$Y_{n,k}(p) = \frac{1}{nh_d}\sum_{i=1}^n K_d\left\{\frac{F_x(g_i) - p}{h_d}\right\} Z_{n,k}(g_i). \tag{A.13}$$

A straightforward but tedious calculation shows that

$$E[Z_{n,1}(p)] = \frac{1}{n}h_r^2\,\mu_2(K_r)\frac{1}{f_X(x)}\{\partial_1^2 h(x,p) - f_X''(x)\,F_x(p)\} + o\left(\frac{h_r^2}{n}\right),$$

$$E[Z_{n,1}(p_1)\,Z_{n,1}(p_2)] = \frac{1}{n^2 h_r}\int K_r^2(u)\,\mathrm{d}u\frac{F_x(p_1\wedge p_2) - F_x(p_1)\,F_x(p_2)}{f_X^2(x)} + o\left(\frac{1}{n^2 h_r}\right),$$

and it follows that

$$E[H_x^{\text{lin}}(p)] = \tilde{b}_n(x,p) + o(h_r^2) + O\left(\frac{1}{nh_d}\right),$$

$$\text{var}\{H_x^{\text{lin}}(p)\} = \frac{1}{nh_r}\tilde{V}(x,p) + o\left(\frac{1}{nh_r}\right).$$

The Lindeberg condition for $Y_{n,1}(p),\ldots,Y_{n,n}(p)$ can easily be checked and the weak convergence in result (A.9) follows.

For a proof of the estimate (A.10) we use Cauchy's inequality, which yields

$$|\Delta_n^{(2)}(p)| \leqslant \frac{1}{nh_d^2}\sum_{i=1}^n \left|K_d'\left(\frac{\xi_i - p}{h_d}\right)\right| D_n^2(g_i)\{1 + o_p(1)\} \leqslant A_n B_n\{1 + o_p(1)\},$$

where the quantities A_n and B_n are defined by

$$A_n = \left\{\frac{1}{nh_d}\sum_{i=1}^n \left|K_d'\left(\frac{\xi_i - p}{h_d}\right)\right|^2\right\}^{1/2},$$

$$B_n = \left\{\frac{1}{nh_d^3}\sum_{i=1}^n D_n^4(g_i)\right\}^{1/2}$$

and $D_n(g_i) = \hat{F}_x(g_i) - F_x(g_i)$. From Lebesgue's theorem and the Lipschitz continuity of the kernel K_d we have for some constant L

$$\begin{aligned} A_n^2 &\leqslant \frac{2}{nh_d}\sum_{i=1}^n \left|K_d'\left\{\frac{F_x(g_i) - p}{h_d}\right\}\right|^2 + \frac{2}{nh_d}\sum_{i=1}^n \left|K_d'\left(\frac{\xi_i - p}{h_d}\right) - K_d'\left\{\frac{F_x(g_i) - p}{h_d}\right\}\right|^2 \\ &\leqslant \frac{2}{h_d}\int_0^1 \left|K_d'\left[\frac{F_x\{G^{-1}(u)\} - p}{h_d}\right]\right|^2 \mathrm{d}u + O\left(\frac{1}{nh_d^2}\right) + \frac{2}{nh_d}\sum_{i=1}^n L\left|\left\{\frac{\xi_i - p}{h_d} - \frac{F_x(g_i) - p}{h_d}\right\}\right|^2 \\ &= 2\int |K_d'(u)|^2\,\mathrm{d}u\frac{G'\{F_x^{-1}(p)\}}{F_x'\{F_x^{-1}(p)\}} + \frac{1}{h_d^2}O_p\left(\frac{1}{nh_r}\right) + o(1) = O_p(1). \end{aligned}$$

A direct calculation yields $\sup_y\{E[D_n^4(x,y)]\} = O(h_r^8)$ and from Markov's inequality we obtain

$$\sqrt{(nh_r)}\,\Delta_n^2(p) = O_p\left(\frac{nh_r^9}{h_d^3}\right)^{1/2}\{1 + o_p(1)\},$$

which proves the remaining statement (A.10) and completes the proof of the first part of theorem 1.

References

Bachrach, L. K., Hastie, T., Wang, M. C., Narashimhan, B. and Marcus, R. (1999) Bone mineral acquisition in healthy asian, hispanic, black and caucasian youth, a longitudinal study. *J. Clin. Endcrin. Metablsm*, **84**, 4702–4712.

Bennett, C. and Sharpley, R. (1988) *Interpolation of Operators*. New York: Academic Press.

Birke, M. and Dette, H. (2007) Estimating a convex function in nonparametric regression. *Scand. J. Statist.*, **34**, 384–404.

Chaudhuri, P. (1991) Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.*, **19**, 760–777.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2007a) Improving estimates of monotone functions by rearrangement. (Available from `arXiv:0704.3686v1`.)

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2007b) Quantile and probability curves without crossing. (Available from `arXiv:0704.3649v1`.)

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

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Fan, J., Yao, Q. and Tong, H. (1996) Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, 189–206.

Gu, C. (1995) Smoothing spline density estimation: conditional distribution. *Statist. Sin.*, **5**, 709–726.

Hall, P., Wolff, R. C. L. and Yao, Q. (1999) Methods for estimating a conditional distribution function. *J. Am. Statist. Ass.*, **94**, 154–163.

Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The Elements of Statistical Learning*. New York: Springer.

He, X. (1997) Quantile curves without crossing. *Am. Statistn*, **51**, 186–191.

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

Koenker, R. (2007) *quantreg: Quantile Regression, R Package Version 4.08*. (Available from `http://www.r-project.org`.)

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

Koenker, R., Ng, P. and Portnoy, S. (1994) Quantile smoothing splines. *Biometrika*, **81**, 673–680.

Koenker, R., Portnoy, S. and Ng, P. (1992) Nonparametric estimation of conditional quantile functions. In *L_1-statistical Analysis and Related Methods* (ed. Y. Dodge), pp. 217–229. Amsterdam: Elsevier.

Masry, E. and Fan, J. (1997) Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.*, **24**, 165–179.

Ng, P. T. and Maechler, M. (2006) *cobs: COBS—Constrained B-splines (Sparse Matrix based), R Package Version 1.1-3.5*. (Available from `http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:cobs`.)

R Development Core Team (2006) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995) An effective bandwidth selector for local least squares regression. *J. Am. Statist. Ass.*, **90**, 1257–1270.

Takeuchi, I., Le, Q. V., Sears, T. D. and Smola, A. J. (2006) Nonparametric quantile estimation. *J. Mach. Learn. Res.*, **7**, 1231–1264.

Volgushev, S. (2007) Nichtparametrische Quantilsregression. *Diplomarbeit*. Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum.

Wand, M. and Ripley, B. (2005) KernSmooth: functions for kernel smoothing for Wand & Jones (1995). In *R Package Version 2.22-17*. (Available from `http://www.maths.unsw.edu.au/wand`.)

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

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

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