

Non-crossing nonparametric estimates of quantile curves

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

Stanislav Volgushev

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: stanislav.volgushev@rub.de

May 8, 2007

Abstract

In this paper a new nonparametric estimate of conditional quantiles is proposed, that avoids the problem of crossing quantile curves [calculated for various $p \in (0, 1)$]. The method uses an initial estimate of the conditional distribution function in a 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 normal distributed and asymptotically first order equivalent to quantile estimates obtained by local constant or local linear smoothing of the 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 proposed method are presented.

Keyword and Phrases: Quantile estimation, conditional distribution, local linear estimate, Nadaraya Watson estimate, crossing quantile curves

1 Introduction

Quantile regression was introduced by Koenker and Bassett (1978) as a supplement to least squares methods focussing on the estimation of the conditional mean function. The very general technique of estimating families of conditional quantile curves yields to a great extension of parametric and

nonparametric 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 modeling. For more details we refer the interested reader to Yu, Lu and Stauder (2003) and the recent monograph of Koenker (2005).

In this article we consider the problem of nonparametric estimation of conditional quantile functions. Since the seminal work of Koenker and Bassett (1978), who considered parametric quantile estimates, several authors have proposed nonparametric methods in this context [see Chaudhuri (1991), Koenker, Portney and Ng (1992, 1994), Yu and Jones (1997, 1998) among many others]. One very popular approach is to combine the concept of local constant or local 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 p th conditional quantile of the conditional distribution $F(y|x)$ is defined as

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

while 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

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

[see for example Yu and Jones (1998) or Koenker (2005), Chap. 7]. In (1.1) and (1.2)

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

denotes the check function, K is a kernel function and h a bandwidth converging to 0 with increasing sample size. As an alternative several authors propose to estimate the conditional distribution function nonparametrically

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

where $w_i(x)$ denotes a weight function depending on x and the predictors $\{X_i\}_{i=1}^n$, and the estimate of the quantile is obtained solving the equation $p = \hat{F}(y|x)$. Note that 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, Wolff and Yao (1999), who proposed adjusted Nadaraya Watson weights].

It was pointed out by several authors [see e.g. He (1997), Yu, Lu and Stauder (2003) or Koenker (2005), Chap. 7] that the estimates 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 to 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. He (1997) proposed a restricted version of regression quantiles in order to avoid the embarrassing phenomenon of quantile crossing, while Yu and Jones (1998) suggested a double kernel smoothing method and in a 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 weights of the Nadaraya-Watson estimate 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 are advantageous 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 the methods (1.1) and (1.2) based on the check function but at the cost of obtaining crossing quantile curve estimates, see Koenker (2005), Chap. 7. Recently Takeuchi, Le, Sears and Smola (2006) proposed to minimize the empirical risk plus a regularizer in order to obtain a nonparametric estimate of the quantile function. As pointed out in this reference the method can be modified to avoid quantile crossing yielding to 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 nonparametric non-crossing estimate of quantile curves. The method is based on an initial nonparametric estimate of the conditional distribution function of the form (1.3) and solves the problem of determining the solution $\bar{q}_p(x)$ of the equation $p = \hat{F}(y|x)$ and monotonicizing the function $p \rightarrow \bar{q}_p(x)$ with respect to the probability p simultaneously. The concept is applicable to any estimate of the conditional distribution function and based on the idea of non-decreasing rearrangements [see Bennett and Sharpely (1988)], which was recently used by Dette, Neumeyer and Pilz (2006) in the context of estimating a monotone regression. The basic arguments of this technique are carefully described in Section 2. While the general methodology is applicable to any estimate of the conditional distribution function, we concentrate in Section 3 on the case of independent

identically distributed observations $\{(X_i, Y_i) \mid i = 1, \dots, n\}$ and on the local constant and local linear estimate of the conditional distribution function [see Fan, Yao and Tong (1996)] for the sake of definiteness. It is demonstrated that the proposed estimates of the conditional quantile curves are asymptotically normal distributed and first order asymptotically equivalent to the estimates proposed by Yu and Jones (1998) and by Hall et al. (1999). We also briefly discuss several extension of the methodology to high dimensional predictors, dependent data and alternative preliminary estimates. In Section 4 some simulation results are presented and the procedure is illustrated by means of a data example. 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 proposed in this paper. Finally, all technical proofs are deferred to an Appendix.

2 Non-crossing nonparametric 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 the sake of simplicity we assume that the explanatory variables are real valued, but the extension to d -dimensional ($d > 1$) covariates will be straightforward [see Remark 3.4 below]. Throughout this paper

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

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

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

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

$$(2.2) \quad H^{-1}(p) := \int_{\mathbb{R}} I\{F(y) \leq p\} dG(y) = \int_0^1 I\{F(G^{-1}(v)) \leq p\} dv.$$

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

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

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

$$(2.4) \quad H_{h_d}^{-1}(p) = \frac{1}{h_d} \int_{-\infty}^p \int_0^1 K_d\left(\frac{F(G^{-1}(v)) - u}{h_d}\right) dv du \xrightarrow{h_d \rightarrow 0} H^{-1}(p),$$

where K_d is a symmetric density (with properties 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 of this fact). Note that the function $H_{h_d}^{-1}$ is always increasing because $(H_{h_d}^{-1})'(p) \geq 0 \forall p \in F(\mathbb{R})$ and that $G^{-1} \circ H_{h_d}^{-1}$ 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 dv in (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

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

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

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

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

Remark 2.1. Note that 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 as local polynomials or smoothing spline methods [see Masry and Fan (1997), Gu (1997)]. Note also that 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 nonvisible.

Remark 2.2. Note that definition (2.6) does not require a local linear or Nadraya Watson type estimate of the conditional distribution function. In fact the procedure can be used to obtain

non-crossing quantile curves from any 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.

3 Asymptotic results

For the sake of definiteness we consider a sample of independent identically distributed observations $(X_1, Y_1), (X_2, Y_2), \dots$, but the results can easily be extended to sequences of dependent random variables [see Remark 3.4 below] As an estimate of the conditional distribution function we use the statistic $\hat{F}_x(y) = \hat{F}(y|x)$ defined in (1.3), where the weights $w_i(x)$ are either the Nadaraya Watson

$$(3.1) \quad w_i(x) = \frac{K_r\left(\frac{X_i-x}{h_r}\right)}{\sum_{j=1}^n K_r\left(\frac{X_j-x}{h_r}\right)}, \quad i = 1, \dots, n,$$

or local linear weights

$$(3.2) \quad w_i(x) = \frac{K_r\left(\frac{X_i-x}{h_r}\right)(S_{n,2} - (x - X_i)S_{n,1})}{S_{n,2}S_{n,0} - S_{n,1}^2}, \quad i = 1, \dots, n,$$

with

$$(3.3) \quad S_{n,\ell} = \sum_{j=1}^n K\left(\frac{x - X_j}{h_r}\right)(x - X_j)^\ell \quad \ell = 0, 1, 2$$

[see Fan and Gijbels (1996)]. In (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 converging to 0 with increasing sample size. The kernel K_d in the statistic is assumed to be twice continuously differentiable, symmetric, positive with compact support $[-1, 1]$, while the corresponding bandwidth h_d converges to 0 with increasing sample size. Furthermore, we assume for the bandwidths h_d and h_r

$$(3.4) \quad nh_r^5 = c + o(1); \quad h_d = o(h_r); \quad \frac{h_r^4}{h_d^3} = o(1)$$

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 obtained from (2.6) with (1.3)

as initial estimate. The proof is deferred to the Appendix. Throughout this paper $\partial_1^k s$ and $\partial_2^k s$ denotes the k th partial derivative of the function $s(x, y)$ or $s(y|x)$ with respect to the coordinate x and y , respectively, and for a kernel K we define $\mu_2(K) = \frac{1}{2} \int_{-1}^1 t^2 K(t) dt$.

Theorem 3.1. *Assume that the functions $f_{(X,Y)}, F_x, G$ are twice continuously differentiable, that the function*

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

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

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

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

$$(3.7) \quad \int |f_X''(u)| du < \infty \quad \int \int |\partial_1 f_{(X,Y)}(u, v)| dudv < \infty$$

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

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

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

with bias

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

and asymptotic variance

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

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

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

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 (3.9).

Remark 3.2. Note that the estimate considered in the second part of Theorem 3.1 is first order asymptotic equivalent to the local double kernel quantile estimate 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 this reference). On the other hand for the local constant and local linear quantile estimate defined by (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 are given by

$$h_r^2 \mu_2(K_r) \left\{ -\frac{\partial_1^2 F(F_x^{-1}(p)|x)}{F'_x(F_x^{-1}(p))} + 2\frac{f'_X(x)(F_x^{-1}(p))'}{f_X(x)} \right\} + o(h_r^2),$$

$$\frac{p(1-p)}{nh_r(F'_x(F_x^{-1}(p)))^2} \int K_r^2(u) du + o\left(\frac{1}{nh_r}\right),$$

respectively. Similarly, the local linear estimate defined by (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 from a first order asymptotic point of view there is a difference between the estimates proposed in this paper and the methods (1.1) and (1.2) based on a combination of the concept of check function and local smoothing. However, the important difference for practical applications is that the new estimates keep the monotonicity with respect to the probability p without any additional modification.

Remark 3.3. It is worthwhile to mention that the inverse of the statistic $\hat{F}_{x,G}^{-1}$, 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 $0 \leq \hat{F}_x(y) \leq 1$ ($y \in \mathbb{R}$). In this sense the methodology can be viewed as an alternative to the adjusted Nadaraya-Watson estimate proposed 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 as presented in the proof of Theorem 3.2 in Dette, Neumeyer and Pilz (2006) that

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

where

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

and

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

This shows that the estimate $\hat{F}_{x,G}$ is first order asymptotically equivalent to the adjusted Nadaraya-Watson estimate proposed by Hall et al. (1999). It is also worthwhile to mention that in contrast to the method suggested by these authors the estimate proposed in this paper can be easily extended to local polynomial techniques [see Fan and Gijbels (1996) or Masry and Fan (1997)] without losing monotonicity and the properties of a distribution function.

Remark 3.4. It follows from the proof in the Appendix that the results of Theorem 3.1 can be generalized in several ways, and some of these generalizations are briefly indicated in this remark.

- (a) Note that a d -dimensional predictor, say $x = (x_1, \dots, x_d)$ does not change the principal result. For example, if the local linear weights together with the corresponding product kernel $K(u_1, \dots, u_d) = \prod_{j=1}^d K_r(u_j)$ are used in the estimate of the conditional distribution function in the first step, the resulting estimate $\hat{F}_{x,G}$ is still asymptotically normal with bias

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

(here the ∂_j^2 denotes the second derivative of $F(y|x)$ with respect to x_j) and variance

$$V(x, p) = \frac{p(1-p) \int K^2(u) du}{f_X(x) n h_r^d (F'_x(F_x^{-1}(p)))^2}.$$

- (b) Results for dependent data can be obtained in a similar way. For example, if $\{(X_i, Y_i)\}$ is an absolute regular process with coefficients

$$\beta(i) = \sup_{i \geq 1} E \left\{ \sup_{A \in \mathcal{F}_1^i} P(A|\mathcal{F}_1^i) - P(A) \right\}$$

satisfying $\sum_{i \geq 1} i^2 \beta(i) < \infty$, then the assertions of Theorem 3.1 are still valid [here \mathcal{F}_j^i denote the sigma field generated by $\{(X_j, Y_j), \dots, (X_i, Y_i)\}$].

Remark 3.5. As indicated in the introduction the procedure proposed in Section 2 is applicable to any nonparametric estimate of the conditional distribution function and similar results as presented in Theorem 3.1 can be obtained provided that there exists a limit theorem for the nonparametric initial estimate of the conditional distribution function, where the bias and variance in Theorem 3.1 have to be modified appropriately. In fact one can (under some very 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 defined in Section 2, say $\hat{F}_x^{-1}(p)$, converges in law, i.e.:

$$a_n \left(\hat{F}_x^{-1}(p) - F_x^{-1}(p) + \frac{b_n(F_x^{-1}(p)|x)}{F'_x(F_x^{-1}(p))} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{\sigma^2(F_x^{-1}(p)|x)}{(F'_x(F_x^{-1}(p)))^2} \right).$$

An important example is the local polynomial estimator of odd degree r [see. Fan, Gijbels (1996) for details]. If the corresponding weights are used for the estimation of the conditional distribution function we obtain

$$\sqrt{nh_r} \left(\hat{F}_x^{-1}(p) - F_x^{-1}(p) + \tilde{b}_n(p|x) \right) \xrightarrow{\mathcal{D}} \mathcal{N} (0, \tilde{\sigma}^2(p|x))$$

with asymptotic bias and variance

$$\begin{aligned} \tilde{b}_n(p|x) &= e_1^t S^{-1} c_r \frac{1}{(r+1)!} \frac{\partial_1^r F(F_x^{-1}(p)|x)}{F'_x(F_x^{-1}(p))} h^{r+1}, \\ \tilde{\sigma}^2(p|x) &= e_1^t S^{-1} S^* S^{-1} e_1 \frac{p(1-p)}{f_X(x)(F'_x(F_x^{-1}(p)))^2}, \end{aligned}$$

where we use the notation

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

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

4 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 impact of the choice of the function G on the estimate of the quantile estimate, secondly we compare the method proposed in this paper with some of the currently available procedures for non-crossing quantile curves, which are most similar in spirit to the new procedure [see Yu and Jones (1998) and Hall et al. (1999)]. Finally the properties of the new estimate are illustrated by a data example.

4.1 Sensitivity with respect to choice of G

Recall from the asymptotic results presented in Section 3 that for the choice $h_d = o(h_r)$ the function G has no impact on the first order 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.

Note that the function G should depend on the point x where one wants 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, \dots, 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 a first step we define for fixed x the set

$$(4.1) \quad \{Y_j \mid |X_j - x| < h_d; \quad j = 1, \dots, n\}$$

of all observations Y_j with a corresponding predictor in the neighbourhood of x , and denote its ordered elements by Z_1, \dots, 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 $\hat{G}_x^{-1}(0.95) = Z_{\lfloor 0.95k \rfloor}$. In order to study the impact of the form of G on the resulting estimate we investigate 3 distributions, the Cauchy, double exponential and normal distribution, where the scaling and location parameter have been adapted to the data by the procedure described in the previous paragraph. In Figure 4.1 we show the three estimates (obtained from the different choices for the function G) for Examples 2 and 3 in Yu and Jones (1998), that is

$$(4.2) \quad \begin{array}{ll} \text{model 2:} & Y = 2.5 + \sin(2X) + 2 \exp(-16X^2) + 0.5Z, \quad X \sim \mathcal{N}(0, 1), Z \sim \mathcal{N}(0, 1) \\ \text{model 3:} & Y = 2 + 2 \cos(X) + \exp(-4X^2) + E \quad X \sim \mathcal{N}(0, 1), E \sim \text{Exp}(1). \end{array}$$

The figure 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 applied the bandwidth selection rule proposed in Yu and Jones (1998)

$$(4.3) \quad h_r = h_r(p) = \hat{h}_{r,opt} \left\{ \frac{p(1-p)}{\varphi(\Phi^{-1}(p))^2} \right\}^{1/5},$$

where φ and Φ denote the density and *cdf* of the standard normal distribution, respectively, and we used the method of Ruppert, Sheather and Wand (1995) to obtain the preliminary estimate $\hat{h}_{r,opt}$. All simulations were done in the statistical software R [R development core team (2006)] and we used the Package 'Kern Smooth' [Wand, Ripley (2005)] to compute the bandwidth. The kernel K_r was chosen as Gauss kernel, while the Epanechnikov kernel was used for K_d . The sample size was $n = 100$, we used $N = 1000$ points to evaluate the function \hat{G}_x^{-1} in (2.5). Note that too large or too small values of the estimate $\hat{F}_{x,G}^{-1}(p)$ would yield to a numerical 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 $0.95I\{\hat{H}_x^{-1}(p) > 0.95\} + \hat{H}_x^{-1}(p)I\{0.05 < \hat{H}_x^{-1}(p) < 0.95\} + 0.05I\{\hat{H}_x^{-1}(p) < 0.05\}$ instead of $\hat{H}_x^{-1}(p)$. The choice of the bandwidth h_d is less critical and we used

$$(4.4) \quad h_d = \min\{p, 1 - p, h_r^{1.3}\}.$$

The results are displayed in Figure 4.1 and in both models we observe no visible differences between the three estimates obtained by different choices for the function G . Further simulation results investigating the impact of the choice of the function G are available from the second author and demonstrate that the results presented in Figure 4.1 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 described at the beginning of this section or according to a similar method, the particular form of the function G has no visible impact on the estimate of the quantile curve for realistic sample sizes .

4.2 Finite sample comparison

We now compare the new estimate with the procedures proposed by Yu and Jones (1998) and Hall et al. (1999), which are most similar in spirit to the method suggested in this paper. For the sake of brevity we restrict ourselves to the four models considered by Yu and Jones (1998). Thus we consider the models 2 and 3 in (4.2) and

$$(4.5) \quad \begin{array}{ll} \text{model 1:} & Y = 1 + \sin(\frac{3}{4}X) + 0.3Z\sqrt{\sin(\frac{3}{4}X) + 1} \quad X \sim \mathcal{N}(0, \frac{1}{16}), Z \sim \mathcal{N}(0, 1) \\ \text{model 4:} & Y = 2 + X + \exp(-X)(E - \log(2.6)) \quad X \sim \mathcal{U}[0; 5], E \sim \text{Exp}(1). \end{array}$$

In Figure 4.2 and 4.3 we display the mean squared error of the estimates of Yu and Jones (1998) (dotted line), Hall et al. (1999) (dashed line) and the method proposed in this paper (solid line). All estimates require a bandwidth for the regression step, which was chosen by (4.3). The method of Yu and Jones (1998) requires a further bandwidth, which was chosen according to formula (12) in that paper, while the bandwidth h_d in the new procedure was given by (4.4) and the function G by the *cdf* of a normal distribution. We display the (pointwise) median of the mean squared error obtained from 1000 simulation runs for the quantiles $p = 10\%$, $p = 50\%$ and $p = 90\%$. In Figure 4.2 we show the results corresponding to model 1 (upper line) and model 2 (lower line). We observe that in model 1 the estimate of Yu and Jones (1998) and the new method yield very similar results for all three quantiles, while the estimate of Hall et al. (1999) has a larger *mse*,

in particular at the boundary of the design space. In model 2 all three estimates have a similar behaviour, where the new method yields a slightly smaller mean squared error if 10% and 90% quantiles have to be estimated. In model 4 [see the lower panel of Figure (4.3)] the situation is slightly different. For the estimation of the 90%-quantile curve all estimates behave similar, while for the 50%-quantile curve the estimate of Yu and Jones (1998) and the procedure proposed in this paper have some advantages compared to method proposed in Hall et. al. (1999). For 10%-quantile curve this method yields better results in the interior than the new procedure, but has some problems in regions closer to the boundary. Here the estimate of Yu and Jones (1998) has the worst performance. In model 3 the estimate of Hall et. al. (1999) yields a larger *mse* when estimating the 10 % and 50% conditional quantiles while the estimate of Yu, Jones (1998) and the new method show a very similar performance. The situation is quite different when estimating the 90% quantile curve, here estimate of Yu, Jones (1998) yields the largest *mse* while the estimator of Hall et. al. (1999) has some advantage over the new method near the boundary. In the interior part of the design region the new method yields a smaller *mse*. Summarizing these results (and further simulations, which are not displayed for the sake of brevity) 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 proposed in this paper. These two estimates show a very similar behaviour with some advantages for the new method.

4.3 Data example

In this section we briefly discuss a data example, which was recently analyzed by Takeuchi et. al. (2006) in order to demonstrate the logical problems arising in crossing estimates of quantile curves. These authors analyzed actual measurements of the relative change in bone mineral density (BMD) at the spine in adolescents as a function of time. The data were originally reported in Bachrach et al. (1999) and were also analyzed in Hastie, Tibshirani, and Friedman (2001) by fitting a smoothing spline. Because the density of the predictors in this example varies substantially, we used a local bandwidth proposed by Fan and Gijbels (1996) for $\hat{h}_{r,opt}$ in equation (4.3). To compute the bandwidth we used the C-code provided by Fan (1995).

In contrast to the analysis performed by Takeuchi et. al. (2006) we consider male and female persons separately [see also Figure 5.6 in Hastie et. al. (1999)]. The corresponding (non-crossing) estimates of the quantile curves are presented in Figure 4.4 and complement the picture observed in the cited references. The relative change in bone mineral density is larger at the beginning of the period between 10 and 25 years and the strong increase happens even earlier for female

persons. The two figures also indicate a larger variability for the female children in regions, where the relative change is rather large.

5 Appendix: proofs

For the sake of brevity we restrict ourselves to a proof of the first part of Theorem 3.1, all other results are proved similarly and the arguments therefore omitted. Moreover, for the sake of a simple notation we assume $n = N$. The proof is performed in two steps. In a first step we establish the weak convergence

$$(4.1) \quad \sqrt{nh_r}(\hat{H}_x^{-1}(p) - H_x^{-1}(p) + \tilde{b}_n(x, p)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{V}(x, p)) ,$$

where

$$(4.2) \quad \tilde{b}_n(x, p) = \frac{h_r^2 \mu_2(K_r)}{f_X(x)} (\partial_1^2 h(x, F_x^{-1}(p)) - f_X''(x)p) \frac{G'(F_x^{-1}(p))}{F_x'(F_x^{-1}(p))} ,$$

$$(4.3) \quad \tilde{V}(x, p) = \frac{p(1-p) \int K^2(u) du (G'(F_x^{-1}(p)))^2}{f_X(x) (F_x'(F_x^{-1}(p)))^2} ,$$

and we use the notation $H_x = F_x \circ G^{-1}$. In a second step we apply a Taylor expansion to complete the proof of Theorem 3.1. 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

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

for some ξ satisfying $|\xi - H_x^{-1}(p)| \leq |\hat{H}_x^{-1}(p) - H_x^{-1}(p)| \xrightarrow{P} 0$. A standard argument and (4.1) now show that $R_n = o_p(1)$, and the assertion of the theorem follows if the first step in (4.1) can be established.

For a proof of the weak convergence in (4.1) we introduce the decomposition

$$(4.4) \quad \hat{H}_x^{-1}(p) - H_x^{-1}(p) = \Delta_n^{(1)}(p) + \frac{1}{2} \Delta_n^{(2)}(p) + \Delta_n^{(3)}(p),$$

where

$$(4.5) \quad \Delta_n^{(1)}(p) = -\frac{1}{nh_d} \sum_{i=1}^n K_d \left(\frac{F_x(g_i) - p}{h_d} \right) (\hat{F}_x(g_i) - F_x(g_i)) ,$$

$$(4.6) \quad \Delta_n^{(2)}(p) = -\frac{1}{nh_d^2} \sum_{i=1}^n K_d' \left(\frac{\xi_i - p}{h_d} \right) \left(\hat{F}_x(g_i) - F_x(g_i) \right)^2,$$

$$(4.7) \quad \Delta_n^{(3)}(p) = \frac{1}{nh_d} \int_{-\infty}^t \sum_{i=1}^n K_d \left(\frac{H(\frac{i}{n}) - u}{h_d} \right) du - H_x^{-1}(p),$$

and we have used the notation $g_i := G^{-1}(i/n)$ and $|\xi_i - F_x(g_i)| \leq |F_x(g_i) - \hat{F}_x(g_i)|$ ($i = 1, \dots, n$)
From Lemma 2.1 in Dette, Neumeier and Pilz (2006) it follows that

$$(4.8) \quad \Delta_n^{(3)}(y) = \mu_2(K_d)h_d^2(H^{-1})''(t) + O\left(\frac{1}{nh_d}\right),$$

and the assertion (4.1) is now a consequence of Slutsky's Lemma if we establish

$$(4.9) \quad \sqrt{nh_r} \left(\Delta_n^{(1)}(p) + \tilde{b}_n(x, p) \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \tilde{V}(x, p)\right),$$

$$(4.10) \quad \sqrt{nh_r} \Delta_n^{(2)}(p) = o_p(1).$$

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

$$(4.11) \quad \Delta_n^{(1)}(p) = -\frac{1}{nh_d} \sum_{i=1}^n K_d \left(\frac{F_x(g_i) - p}{h_d} \right) D_n(p, g_i)(1 + o_p(1)),$$

$$(4.12) \quad \Delta_n^{(2)}(p) = -\frac{1}{nh_d^2} \sum_{i=1}^n K_d' \left(\frac{\xi_i - p}{h_d} \right) D_n^2(x, g_i)(1 + o_p(1)),$$

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

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

with $Z_{n,i}(y) = \frac{1}{nh_r} \frac{1}{f_X(x)} K_r \left(\frac{X_i - x}{h_r} \right) (I_{\{Y_i \leq y\}} - F_x(y))$. Consequently, the leading term in (4.11) can be written as $H_x^{lin}(p) = \sum_{k=1}^n Y_{n,k}(p)$, where the independent, identically distributed random variables $Y_{n,k}(p)$ are given by

$$(4.13) \quad Y_{n,k}(p) = \frac{1}{nh_d} \sum_{i=1}^n K_d \left(\frac{F_x(g_i) - p}{h_d} \right) Z_{n,k}(g_i).$$

A straightforward but tedious calculation shows that

$$\begin{aligned} \mathbb{E}[Z_{n,1}(p)] &= \frac{1}{n} h_r^2 \mu_2(K_r) \frac{1}{f_X(x)} (\partial_1^2 h(x, p) - f_X''(x) F_x(p)) + o\left(\frac{h_r^2}{n}\right), \\ \mathbb{E}[Z_{n,1}(p_1) Z_{n,1}(p_2)] &= \frac{1}{n^2 h_r} \int K_r^2(u) du \frac{F_x(p_1 \wedge p_2) - F_x(p_1) F_x(p_2)}{f_X^2(x)} + o\left(\frac{1}{n^2 h_r}\right), \end{aligned}$$

and it follows that

$$\begin{aligned} \mathbb{E}[H_x^{lin}(p)] &= \tilde{b}_n(x, p) + o(h_r^2) + O\left(\frac{1}{nh_d}\right) \\ \text{Var}(H_x^{lin}(p)) &= \frac{1}{nh_r} \tilde{V}(x, p) + o\left(\frac{1}{nh_r}\right). \end{aligned}$$

The Lindeberg condition for $Y_{n,1}(p), \dots, Y_{n,n}(p)$ can easily be checked and the weak convergence in (4.9) follows.

For a proof of the estimate (4.10) we use Cauchy's inequality which yields

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

where the quantities A_n and B_n are defined by

$$\begin{aligned} A_n &= \left(\frac{1}{nh_d} \sum_{i=1}^n \left| K'_d \left(\frac{\xi_i - p}{h_d} \right) \right|^2 \right)^{1/2} \\ B_n &= \left(\frac{1}{nh_d^3} \sum_{i=1}^n D_n^4(g_i) \right)^{1/2} \end{aligned}$$

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

$$\begin{aligned} A_n^2 &\leq \frac{2}{nh_d} \sum_{i=1}^n \left| K'_d \left(\frac{F_x(g_i) - p}{h_d} \right) \right|^2 + \frac{2}{nh_d} \sum_{i=1}^n \left| K'_d \left(\frac{\xi_i - p}{h_d} \right) - K'_d \left(\frac{F_x(g_i) - p}{h_d} \right) \right|^2 \\ &\leq \frac{2}{h_d} \int_0^1 \left| K'_d \left(\frac{F_x(G^{-1}(u)) - p}{h_d} \right) \right|^2 du + O\left(\frac{1}{nh_d^2}\right) + \frac{2}{nh_d} \sum_{i=1}^n L \left| \left(\frac{\xi_i - p}{h_d} - \frac{F_x(g_i) - p}{h_d} \right) \right|^2 \\ &= 2 \int |K'_d(u)|^2 du \frac{G'(F_x^{-1}(p))}{F'_x(F_x^{-1}(p))} + \frac{1}{h_d^2} O_p\left(\frac{1}{nh_r}\right) + o(1) = O_p(1). \end{aligned}$$

A direct calculation yields $\sup_y \mathbb{E}[D_n^4(x, y)] = O(h_r^8)$ 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 (4.10) and completes the proof of the first part of Theorem 3.1. □

Acknowledgements. The authors are grateful to Isolde Gottschlich who typed numerous versions of this paper with considerable technical expertise and to R. Koenker, who pointed out that

the problem of crossing estimates of conditional quantile functions is of importance. 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 a NIH grant award IR01GM072876:01A1.

References

- L. K. Bachrach, T. Hastie, M. C. Wang, B. Narashimhan, and R. Marcus (1999). Bone mineral acquisition in healthy asian, hispanic, black and caucasian youth, a longitudinal study. *Journal of Clinical Endocrinal Metabolism* 84, 4702-4712.
- C. Bennett, R. Sharpley (1988). *Interpolation of Operators*. Academic Press, N.Y.
- P. Chaudhuri (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* 19, 760-777.
- H. Dette, N. Neumeier, K.F. Pilz (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* 12, 469-490.
- Fan, J. (1995) <http://orfe.princeton.edu/~jqfan/fan/publications.html>
- Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall, London.
- J. Fan, Q. Yao, H. Tong (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* 83, 189-206.
- C. Gu (1995). Smoothing spline density estimation: conditional distribution. *Stat. Sinica* 5, 709-726.
- P. Hall, R.C.L. Wolff, Q. Yao (1999). Methods for estimating a conditional distribution function. *J. Am. Stat. Assoc.* 94, 154-163.
- T. Hastie, R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning*. Springer, New York.
- X. He (1997). Quantile curves without crossing. *The American Statistician* 51, 186-191.
- R. Koenker, G. Bassett (1978). Regression quantiles. *Econometrica* 46, 33-50.
- R. Koenker, S. Portney, P. Ng (1992). Nonparametric estimation of conditional quantile functions. In: *L₁-Statistical Analysis and Related Methods*, ed. Y. Dodge, Amsterdam: Elsevier, 217-229.

- R. Koenker, S P. Ng, . Portney (1994). Quantile smoothing splines. *Biometrika* 81, 673-680.
- R. Koenker (2005). *Quantile Regression*. Cambridge University Press.
- E. Masry, J. Fan (1997). Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.* 24, 165-179.
- R Development Core Team (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- D. Ruppert, S.J. Sheather, M.P. Wand (1995). An effective bandwidth selector for local least squares regression. *J. Am. Stat. Assoc.* 90, 1257-1270.
- I. Takeuchi, Q.V. Le, T. D. Sears A.J. Smola (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research* 7, 12311264.
- S original by Matt Wand. R port by Brian Ripley. (2005). *KernSmooth: Functions for kernel smoothing for Wand & Jones (1995)*. R package version 2.22-17.
<http://www.maths.unsw.edu.au/wand>
- K. Yu, Z. Lu, J. Stauder (2003). Quantile regression: applications and current research areas. *The Statistician* 52, 331-350.
- K. Yu, M.C. Jones (1997). A comparison of local constant and local linear regression quantile estimators. *Comput. Stat. Data Anal.* 25, 159-166.
- K. Yu, M.C. Jones (1998). Local linear quantile regression. *J. Am. Stat. Assoc.* 93, 228-237.

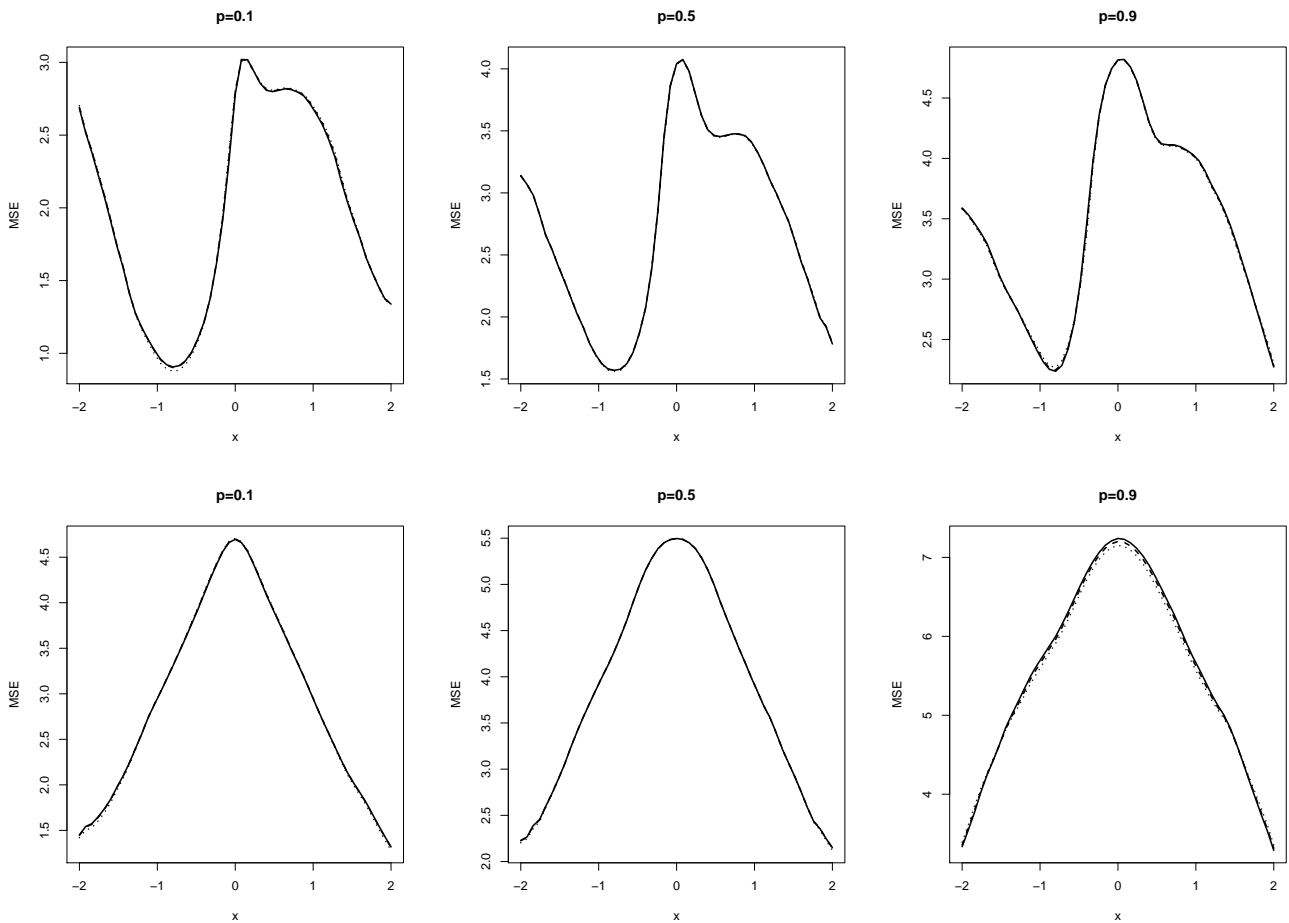


Figure 4.1: *Nonparametric estimates of the quantile curves for various functions G and various quantiles for models 2 (upper panel) and 3 (lower panel). The estimates result from the use of the normal (solid line), double exponential (dashed line), and Cauchy distribution (dotted line).*

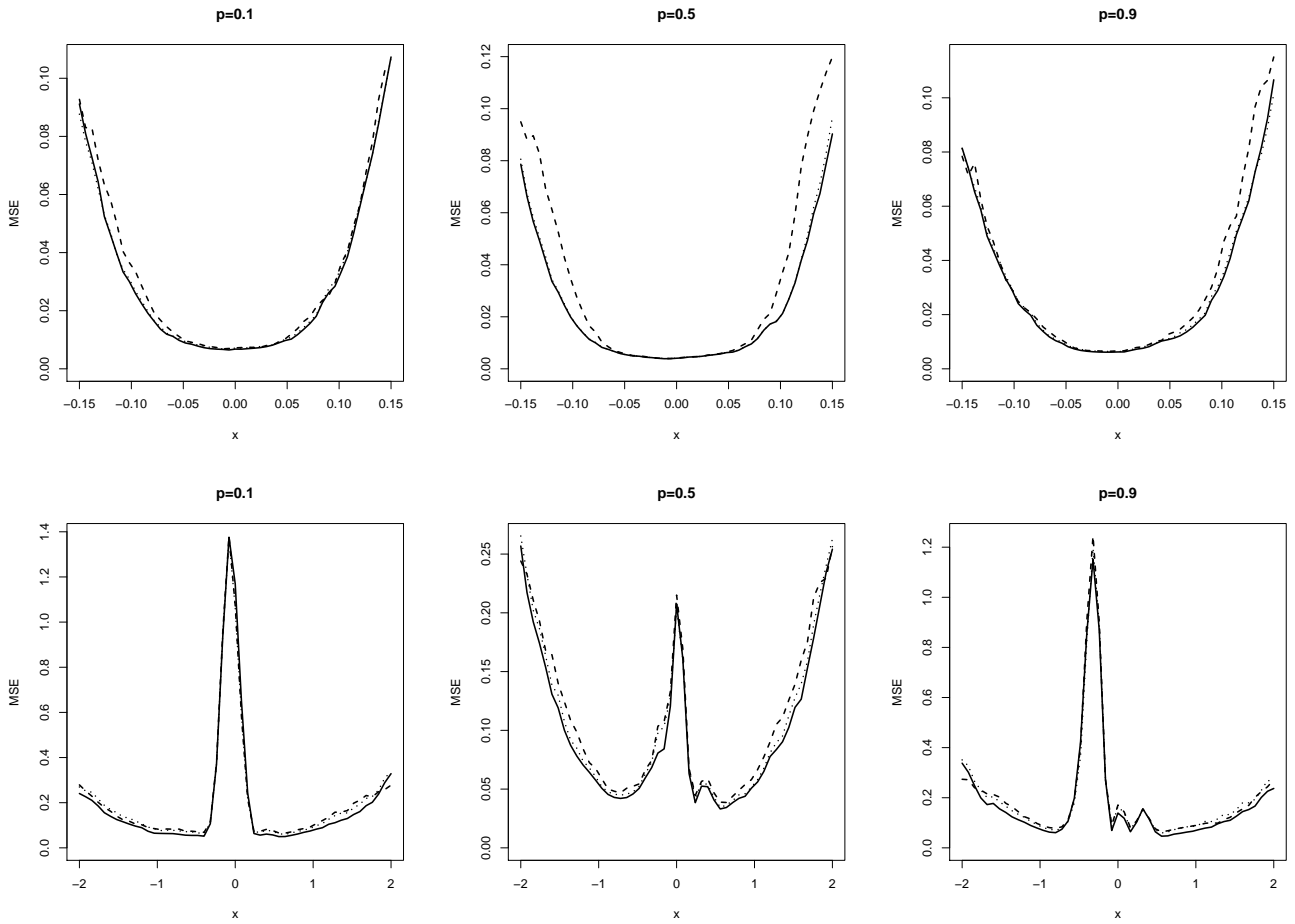


Figure 4.2: Mean squared error of nonparametric estimates in model 1 (upper panel) and model 2 (lower panel) for different quantile estimates. Dashed line: estimate of Hall et al. (1999); dotted line: estimate of Yu and Jones (1998); solid line: the estimate $\hat{F}_{x,G}$ proposed in this paper.

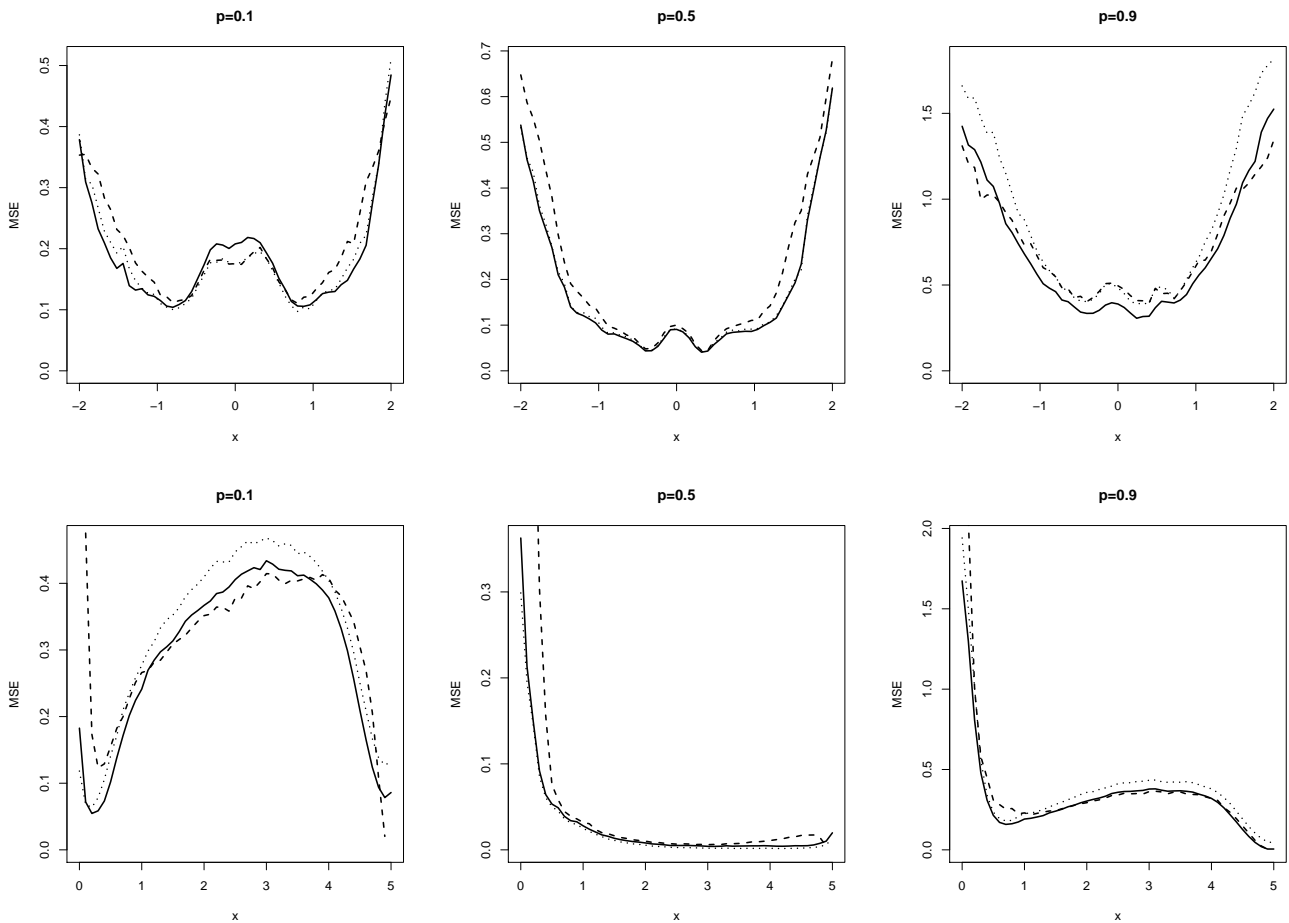


Figure 4.3: Mean squared error of nonparametric estimates in model 3 (upper panel) and model 4 (lower panel) for different quantile estimates. Dashed line: estimate of Hall et al. (1999); dotted line: estimate of Yu and Jones (1998); solid line: the estimate $\hat{F}_{x,G}$ proposed in this paper.

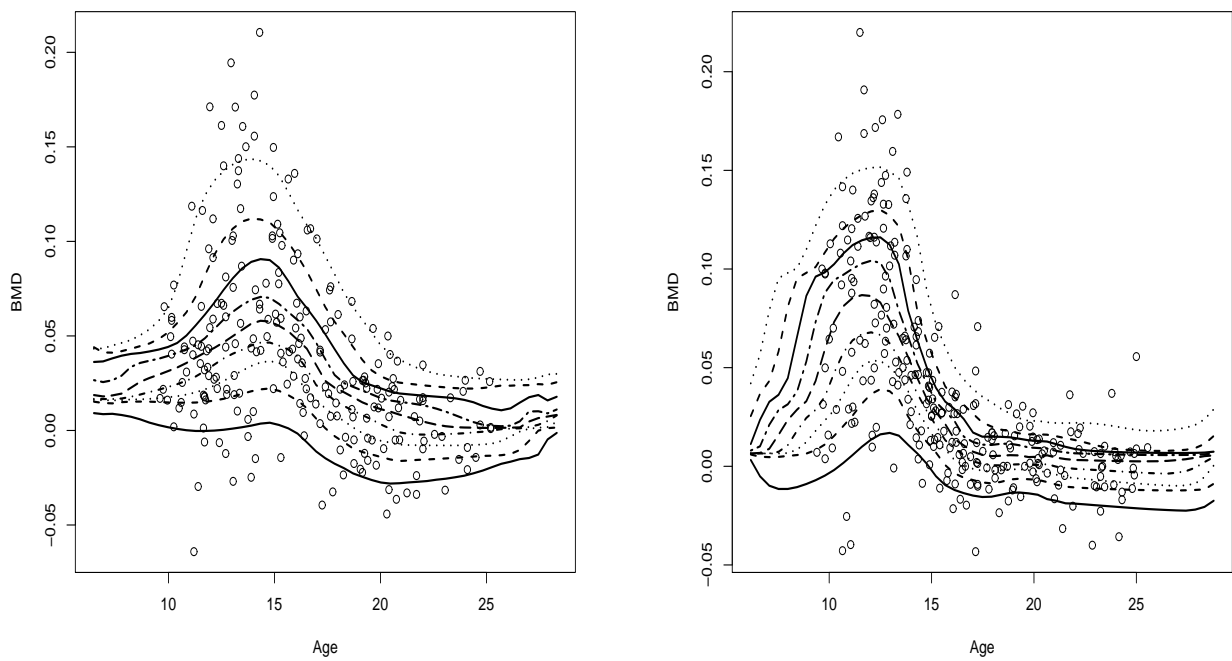


Figure 4.4: *Non-crossing estimates of quantile curves for the BMD data. Left panel: male persons, right panel: female persons.*