UNIVERSITÀ DELLA CALABRIA

**Dipartimento di Economia,
Statistica e Finanza
Ponte Pietro Bucci, Cubo 0/C
87036 Arcavacata di Rende (Cosenza) - Italy**
http://www.unical.it/desf/

Working Paper n. 08 - 2013

WEIGHTED NON-CROSSING QUANTILE REGRESSIONS

Ilaria Lucrezia Amerise
Dipartimento di Economia,
Statistica e Finanza
Università della Calabria
Ponte Pietro Bucci, Cubo 0/C
Tel.: +39 0984 492474
Fax: +39 0984 492421
e-mail: ilaria.amerise@unical.it

Novembre 2013

Weighted non-crossing quantile regressions

Ilaria Lucrezia Amerise*

Dipartimento di Economia, Statistica e Finanza
Via Pietro Bucci, Cubo 0C
87036 Rende (CS) - Italy

e-mail: ilaria.amerise@unical.it

Abstract: In this article we are concerned with a collection of multiple linear regressions that enable the researcher to gain an impression of the entire conditional distribution of a response variable given the specifications for the explanatory variables. In particular, we investigate the advantage of using a new method of parametric estimation for non-crossing quantile regressions. The main tool is a weighting system of the observations that aims to reduce the effect of contamination of the sampled population on the estimated parameters by diminishing the effect of outliers.

The performance of the new estimators has been evaluated on a number of data sets. We had considerable success with avoiding intersections and in the same time improving the global fitting of conditional quantile regressions. We conjecture that in other situations (*e.g.* data with high level of skewness, non-constant variances, unusual and uncertain data) the method of weighted non-crossing quantiles will lead to estimators with good robustness properties.

Keywords and phrases: conditional quantiles, monotonicity problem, estimation under constraints
JEL Classification: C21, C31, C6.

1. Methodology and estimation

A typical investigation in statistical analysis consists of the linear regression of one response variable onto one or more predictor or explanatory variables, where the data are observed on a sample of entities. The rationale is that by establishing a relationship between them, knowledge of the value of predictor variables enables an approximate value to be predicted for the response variable. However, a richer and more precise understanding can be achieved through quantile regression analysis, which allows the researcher to examine and compare different levels of response, given the variation in the explanatory variables, for a properly chosen set of quantiles.

Let $Q_p(Y|\mathbf{x}) = \inf\{Pr(Y \leq y|\mathbf{x}) \geq p\}$ indicate the p-th conditional quantile ($0 < p < 1$) of a real valued random variable Y given a vector of m explanatory variables $\mathbf{x}$. In short, $Q_p(Y|\mathbf{x})$ is the smallest real value such that the probability of obtaining smaller values of Y is at least p. In general, the quantiles are group of values that

*La presente pubblicazione è cofinanziata con il sostegno della Commissione Europea, Fondo Sociale Europeo e della Regione Calabria. L'autore è il solo responsabile di questa pubblicazione, la Commissione Europea e la Regione Calabria declinano ogni responsabilità sull'uso che potrà essere fatto delle informazioni in essa contenute

divide the total probability into parts. Group of interest are obtained by using the median which divides the distribution of $Y|\mathbf{x}$ into halves ($p = 0.5$), the three quartiles which divide the distribution into four equal parts ($p = 0.25h, h = 1, 2, 3$), deciles ($p = 0.1h, h = 1, \cdots, 9$), and so on. Quantiles could be considered also at irregularly spaced over the $(0, 1)$ interval of probabilities.

For a random sample of observations $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ of Y, a linear regression model may be specified as

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta}(p) + e_{i,p} \quad \text{for } i = 1, 2, \cdots, n; \quad p \in (0, 1) \tag{1.1}$$

with $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,m}), i = 1, 2, \cdots, n$ being a sequence of $m \times 1$ vectors of known values and $n > m$. For each p, the magnitude and the direction of the effect of a given predictor can be compared non only with the effect of the other predictors in the same equation, but also with the effect of the same predictor for other quantile equations. The vector $\boldsymbol{\beta}(p) \in R^m$ contains m coefficients whose estimate should be obtained from sample data. The individual coefficient $\hat{\beta}_j(p), j = 1, 2, \cdots, m$ can be interpreted as the trade-off $\frac{\partial Q_p(y|\mathbf{x})}{\partial x_j}$ between a marginal change in the p-th conditional quantile of the response variable and a marginal change in the value of the j-th predictor, when all of the other variables are held constant. If $\boldsymbol{\beta}(p)$ is fixed at $\boldsymbol{\beta}$ for each p, model (1.1) reduces to the standard conditional expectation model with heteroscedastic disturbances $y_i = \mathbf{x}_i^t \boldsymbol{\beta} + e_{i,p}$.

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ denote the $n \times m$ matrix with columns $\mathbf{x}_i, i = 1, 2, \cdots, m$. We assume that the explanatory variables include an intercept term and, therefore, the first column of $\mathbf{X}$ consists entirely of ones. We assume further that $\mathbf{X}$ has rank m and that $e_1, e_2, \cdots, e_n$ are independent random disturbances with quantile function $Q_p(e)$. The quantile function is left unspecified; we only require $Q_p(e_{i,p}|\mathbf{x}_i) = 0$, which implies that the conditional p-th quantile of $e_{i,p}$ is null for each i. It follows that the p-th conditional quantile of $y|\mathbf{x}_i$ is given by

$$Q_p(y_i|\mathbf{x}_i) = Q_p\left[\mathbf{x}_i^t \boldsymbol{\beta}(p) + e_{p,i}|\mathbf{x}_i\right] = Q_p\left[\mathbf{x}_i^t \boldsymbol{\beta}(p)|\mathbf{x}_i\right] + Q_p\left[e_{p,i}|\mathbf{x}_i\right] = \mathbf{x}_i^t \boldsymbol{\beta}(p). \tag{1.2}$$

Note that $Q_p\left[\mathbf{x}_i^t \boldsymbol{\beta}(p)|\mathbf{x}_i\right] = \mathbf{x}_i^t Q_p\left[\boldsymbol{\beta}(p)|\mathbf{x}_i\right] = \mathbf{x}_i^t \boldsymbol{\beta}(p)$ by the definition of the conditional expectation. It is worth noting that there is no assumption on identical distributions and that model (1.1) allows the disturbances to change as a function of $\mathbf{X}$ and, thus, various form of heteroscedasticity and local noise rates can be accommodated.

A quantile regression estimate $\hat{\boldsymbol{\beta}}(p)$ of the unknown parameters is defined as that values of $\boldsymbol{\beta}$ that minimize the asymmetrical loss function

$$\mathbf{Q}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{x}) = \min_{\boldsymbol{\beta} \in R^m} \left\{ p \sum_{i|y_i \geq \mathbf{x}_i^t \boldsymbol{\beta}} \left(y_i - \mathbf{x}_i^t \boldsymbol{\beta}\right) + (1 - p) \sum_{i|y_i < \mathbf{x}_i^t \boldsymbol{\beta}} \left(\mathbf{x}_i^t \boldsymbol{\beta} - y_i\right) \right\}. \tag{1.3}$$

The minimizing $\hat{\boldsymbol{\beta}}(p)$ determine a m-dimensional hyperplane defined as $\mathbf{x}_i^t\hat{\boldsymbol{\beta}}(p)$ that best fits the n observations. All observations above the best interpolating hyperplane contribute with weight p to the estimates of the parameters; all observations below the hyperplane contribute with weigh $(1-p)$.

The intuition behind the seminal article of Koenker & Bassett [1978] is quite simple. There is a complete equivalence between the computation of a quantile in terms of the order statistics $y_{(i)}, i = 1, \cdots, n$ and the minimization of an asymmetrical loss function such as (1.3) which, in turn, can be reformulated as the minimization of a linear function subject to linear constraints.

$$\min_{\boldsymbol{\beta}(p)\in R^m;\mathbf{r},\mathbf{s}\in R^n} \begin{bmatrix} p\mathbf{r}^t\mathbf{u}_n & (1-p)\,\mathbf{s}^t\mathbf{u}_n & \mathbf{0}_n^t \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \\ \boldsymbol{\beta}(p) \end{bmatrix} \tag{1.4}$$

subject to

$$\begin{bmatrix} -\mathbf{I}_{n,n} & \mathbf{I}_{n,n} & -\mathbf{X} \\ \mathbf{I}_{n,n} & -\mathbf{I}_{n,n} & \mathbf{X} \\ \mathbf{I}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{I}_n & \mathbf{0}_{n,n} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \\ \boldsymbol{\beta}(p) \end{bmatrix} \geq \begin{bmatrix} \mathbf{y} \\ \mathbf{y} \\ \mathbf{0}_n \\ \mathbf{0}_n \end{bmatrix} \tag{1.5}$$

where $\mathbf{u}_n$ is the $(n \times 1)$ vector of ones, $\mathbf{I}_{n,n}$ is the identity matrix of order n, $\mathbf{0}_{n,n}$ is the $(n \times n)$ matrix of zeros and $\mathbf{0}_n$ is the $(n \times 1)$ vector of zeros.

The linearity of the objective function and the linear constraints imply that a solution of (1.3) has to lie in one of the vertices of the polyhedron defined in (1.4). The advent of modern linear programming techniques in the later 1940s lead to fast and efficient algorithm to solve this type of problems. In fact, it may be the computational complexity of least absolute regression, as well as the analytical intractability in a statistical setting that forced the least absolutes approach to take the back seat to least squares in multiple linear regression. See Seneta & Steiger [1984]. Here we will not go deep into the computational aspects of quantile regression, but just recall the availability of efficient algorithms (Koenker & D'Orey, 1987, 1994) and Portnoy [1991].

Buchinsky [1998] notes that the $m \times 1$ vector of first-order conditions for solving the problem in (1.3) is given by

$$n^{-1}\sum_{i=1}^{n}\left[p - \frac{1}{2} + \frac{1}{2}sgn\left(y_i - \mathbf{x}_i^t\boldsymbol{\beta}\right)\right]\mathbf{x}_i \;\approx\; 0 \tag{1.6}$$

where the $sgn\,(.)$ (*signum*) function takes the values $-1, 0, +1$ according to whether its argument is negative, zero or positive. The approximation symbol emphasizes the fact that since (1.6) is a discontinuous function of $\boldsymbol{\beta}$, it may not have an exact solution. However, if $n \to \infty$ then (1.6) converges to zero.

Let $M = (r_1, \cdots, r_m)$ be a subset of m distinct integers from $\{1, \cdots, n\}$ defining a combination of rows such that $\mathbf{X}(M)$ has full rank m. According to Koenker & Bassett [1978] (Theorem 3.1) the solutions to (1.3) have the form

$$\begin{aligned} & & [\mathbf{X}(M)]^{-1}\mathbf{y}(M) \\ \mathbf{r} &= \max\{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(p), \mathbf{0}_n\}, & \mathbf{r}_M &= \mathbf{0}_M \\ \mathbf{s} &= -\min\{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(p), \mathbf{0}_n\}, & \mathbf{s}_M &= \mathbf{0}_M \end{aligned} \tag{1.7}$$

where $\mathbf{0}_n$ and $\mathbf{0}_M$ are vectors of n and M zeros, respectively. Moreover, if the quantile function of the disturbances is continuous, then problem (1.3) has a unique solution $\hat{\boldsymbol{\beta}}(p)$ provided that

$$(p-1)\,\mathbf{u}_k^t < \sum_{i \notin r} \left\{0.5\left[1 - sgn\left(y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}}(p)\right)\right] - p\right\} \mathbf{x}_i \mathbf{X}(M)^{-1} < p\mathbf{u}_k^t. \tag{1.8}$$

The residuals of the estimated quantile regression have an interesting structure. Koenker & Bassett [1978] (Theorem 3.4) show that

$$\frac{n^-}{n} \leq p \leq 1 - \frac{n^+}{n} \tag{1.9}$$

where n^- and n^+ indicate, respectively, the number of negative and positive residuals. In the case the solution of (1.3) is unique, all inequalities are strict. Furthermore, when the quantile function of the disturbances is continuous then there are exactly m residuals with value zero. By definition, a percentage p of observed values is less than the fitted values and a percentage $(1-p)$ of the observed values greater than that of the fitted values.

For $n \to \infty$, we could estimate an increasing number of quantile regressions; in practice, there may be at most $3n$ distinct regression solutions for $p \in [0, 1]$ (see Koenker & D'Orey, 1987). In finite samples, Portnoy [1991] shows that the number of distinct quantile regressions is $O(nlogn)$. The common practice is that quantile regression is designed to be used in groups rather than singly where the number of elements in a group may increase as the sample size n increases. This opportunity is particularly useful when the predictors have a different impact on different regions of the design space. For example, pairs of extreme conditional quantiles map out a conditional prediction interval within which one expects a specified fraction of individual points to lie. Also, for unimodal distributions, the analysis of kurtosis can focus on how the covariates affect both the tails and the central parts of the conditional distribution.

Finite sample and asymptotic theory of quantile regression is not discussed in this paper because of the vast literature accumulated on the subject. The survey in Koenker [2005][Ch.3 and Ch. 4] is particularly effective in this sense.

2. Crossing quantile regressions

Quantile regression estimates are robust in presence of observations that are far in the direction of the response variable. This is an attractive property, at least in part derived from ordinal nature of the quantiles, which slows down the leverage from outlying observations. At the other side of the coin, there is the potential drawback that quantile regression estimates are not guaranteed to be unique for the given percentage p.

When a regression model is assessed, the two main characteristics that need to be considered are robustness and sensitivity. Robustness is a valuable characteristic because quantile regression does not change greatly when data are changed slightly. However, since robustness is achieved at the cost of a loss in precision, it can become a problem if the gaps between the percentages are too narrow. Sensitivity is important, but it probably reduces the reliability of estimation when substantially similar observations are mapped onto very distant conditional values of the response. Robustness and sensitivity are antithetical requirements because robust procedures give greater stability against random changes in data, whereas more sensitive procedures offer a richer source of information regarding the dependence structure. A balanced solution may be the analysis of the conditional quantile function for an appropriate set of percentages $p \in P\,(0 < p_1 < \cdots < p_k < 1)$ and the estimation of the parameters separately for each quantile regression. It should nonetheless be kept in mind that, when several conditional quantiles are treated, it is not unusual that the estimated parameters generate non-parallel hyperplanes. When k increases and a limited amount of data is available, the phenomenon of crossing hyperplanes becomes much more likely.

Quantile regression hyperplanes in R^m are defined by

$$\left\{ \mathbf{x} \in R^m \middle| \sum_{j=1}^{m} \beta_j(p)\, x_j - y = 0 \right\} \quad \text{with} \quad \boldsymbol{\beta}(p) \neq \mathbf{0}_m. \tag{2.1}$$

where $\mathbf{0}_m$ is a $(m \times 1)$ vector of zeros. Note that two equations form the same hyperplane if and only if they differ by a multiplicative factor not equal to zero. A violation of the monotonicity condition occurs when for two quantiles p_1, p_2 and a $(m \times 1)$ vector of explanatory variables values $\mathbf{x}$ we have

$$\mathbf{x}^t \boldsymbol{\beta}(p_1) \;=\; \mathbf{x}^t \boldsymbol{\beta}(p_2) \quad \text{for } p_1 \neq p_2, \quad \mathbf{x} \in R^m. \tag{2.2}$$

If the vectors $\boldsymbol{\beta}(p_1)$ and $\boldsymbol{\beta}(p_2)$ are linearly independent, then there are two independent linear equations in m unknowns. After solving the first equation for x_r, this value can be substituted into the second equation, which can be solved for $x_s, s \neq r$. At this point there are $(m-2)$ free unknowns. Any two non-parallel hyperplanes intersect in one hyperplane of dimension $(n-2)$.

To illustrate, consider the quantile regression model (1.1) with $m = 2$. In this case we have $\mathbf{x}_i = (1, x_i)$ and

$$Q_p(y_i|x_i) = \beta_0(p) + \beta_1(p)\, x_i \quad \text{for } i = 1, 2, \cdots, n. \tag{2.3}$$

If the support of x_i is the entire real line, then either $\beta_1(p)$ is a constant independent of p or two or more conditional quantile regressions overlap for some value of x_i, not necessarily in the range of the observed values. This simply implies that $y|\mathbf{x}_i$ is higher at a lower quantile and vice versa. For example, a given point (y, x) might result simultaneously below $p_1 = 0.20$, but above $p_2 = 0.25$ leading to an invalid $y|x$ distribution. He [1997] observes that crossing quantiles hyperplanes reflects a paucity of data in the region concerned (a sort of misspecification of the predictor effects). In this sense, Koenker & Geling [2001] suggest introducing additional predictors to avoid crossing. For example, we can vary the specification of the model for each quantile by adding and subtracting a positive variable

$$Q_p\left(y_i|\mathbf{x}_i\right) = \beta_0(p) + \beta_1(p)\, x_i \pm \exp\left[\beta_2\left(p\right) x_i\right] \quad \text{for } i = 1, 2, \cdots, n. \tag{2.4}$$

This expression incorporates two quantile regressions that never cross one another and do not cross the line (2.3). In the multivariate case, crossing could be avoided if all quantile hyperplanes are parallel. For instance, Zhao [2000] first estimates the slope parameters by the least absolute deviation ($p = 0.5$). Common slopes guarantees that all the quantile hyperplanes will be parallel with no intersection. Second, the estimates of the intercepts are obtained at different quantiles of the residuals determined in the first step. The combined estimates produce a consistent estimator of the theoretical regression quantile. Note that this is the only possible solution when the support of the explanatory variables is the entire R^m. In the same spirit, Todkar & Kadane [2012] build a model of quantile regression monotonically increasing in $p \in [0, 1]$ obtained by reparametrizing the elements of $\boldsymbol{\beta}(p)$ as linear combinations of two monotonically increasing curves.

Bassett & Koenker [1982] (Theorem 2.1) show that the estimated conditional quantile function at the centroid $\mathbf{x} = \bar{\mathbf{x}}$ (the vector whose the i-th element $\bar{x}_i$ is the average of $\mathbf{x}_i$) we have $Q_p\left(y_i|\bar{\mathbf{x}}\right) = \bar{\mathbf{x}}^t\boldsymbol{\beta}(p)$, which is a monotone jump function of p on the interval $[0, 1]$. Moreover, $Q_p\left(y_i|\mathbf{x}\right)$ must be monotonic in p in a neighborhood of $\mathbf{x} = \bar{\mathbf{x}}$. Thus, incidence of crossing generally occur only in the extreme regions of the domain S^m, that is the convex hull of the n data $\mathbf{x} \in R^m$. More specifically, S^m is the intersection of all convex sets containing the observations in $\mathbf{X}$.

$$S^m \equiv \left\{ \sum_{i=1}^{m} \lambda_i \mathbf{x}_i | \lambda_i \geq 0 \ \forall\ i, \quad \sum_{i=1}^{m} \lambda_i = 1 \right\}. \tag{2.5}$$

On the other hand, we should ignore points close to the boundary or lying outside S^m unless the data set include sufficient observation in the extreme regions to allow a

reliable computation of quantiles. Schnabel & Eilers [2013] point out that, although in many cases crossing is only a visual annoyance, it may jeopardize further analysis, *e.g.* when studying conditional distributions at specific values of the independent variable.

Convergence to the true conditional quantile functions renders legitimate the expectation that the crossing phenomenon will eventually disappear as the sample size n increases. Machado & Mata [2005] recall the theoretical results of Bassett & Koenker [1982] (Theorem 3.2) and Bassett & Koenker [1986] (Theorem 3.1), which show that the estimated parameters of the quantile regression are consistent for their population counterpart. The theory, therefore, predicts that the potential violations of monotonicity will be smaller the larger the sample size and (the sparser the set of $p \in [0, 1]$). This is not necessarily true for a general matrix $\mathbf{X}$ and the estimated hyperplanes for the actual data set may overlap. On the other hand, because of the phenomenon known as the "course of dimensionality" (which is virtually omnipresent when analyzing data in high-dimensional spaces) even large datasets may become rarefied in certain regions to a degree which favors quantile crossings.

2.1. Literature review and research directions

Bondell *et al.* [2010] observe that quantile crossing is a well-known problem, but no simple and general solution currently exists. In order to circumvent this difficulty, many authors have looked for techniques that are capable of fitting the data appropriately and several attempts at this have been made since the late 1990s. Literatures on avoid crossings can generally be divided into two major approaches: semi-parametric techniques, where the underlying error quantile function does not assume any specific form and non-parametric methods where various smoothing techniques (*e.g.* kernel fitting or polynomial spline fitting) are adapted to the error distribution.

Two methods to prevent quantile inversions, one for each approach, were proposed by He [1997]: the first, applied the Box-Cox transformation to restrict regression quantiles (on this see also Heagerty & Pepe, 1999). The second imposes certain restrictions on the space of possible solutions to conditional quantiles. The restricted regression quantile curves are not easy to compute, but do not suffer from the problem of quantile reversal in certain areas of the space of the explanatory variables.

Yu & Jones [1998] study nonparametric regression quantile estimation by kernel weighted local linear fitting. Specifically, given the current quantile function, the next quantile function is estimated so that it does not cross with the existing quantiles. The authors show that local linear conditional quantile estimation is feasible

and practical. Results are at the least comparable with those produced by other approaches.

Takeuchi & Furuhashi [2004] addressed the problem following a support vector machine approach. With the use of kernel-based estimator, a non-crossing conditional quantile estimator is derived in the form of a constrained maximization of a piecewise quadratic function. See also Takeuchi *et al.* [2006].

To deal with the potential lack of monotonicity in multiple quantile regressions, Melly [2005] developed a two-step procedure. In the first stage, the model $y_i = \mathbf{x}_i^t \boldsymbol{\beta}(p)$ is estimated along a grid of k different p-values whose mesh is sufficiently dense (a mesh size of order $O\left(n^{-(0.5+\epsilon)}\right)$ will work). In the second stage, quantiles of the $k * n$ estimates $\mathbf{x}_i^t \hat{\boldsymbol{\beta}}(p_h)\,,\ h = 1, \cdots, k; i = 1, \cdots, n$ are computed by weighting each element by $(p_h - p_{h-1})$. The result is the estimation of the unconditional quantiles of y

Neocleous & Portnoy [2007] show that by choosing an appropriate grid of p-values and defining the quantile functions by linear interpolation between grid values, the resulting conditional quantile estimator is strictly monotonic with probability tending to one, and it is asymptotically equivalent to the usual regression quantile estimator.

Dette & Volgushev [2008] proposed non-crossing estimates of quantile curves using a simultaneous inversion and isotonization of an estimate of the conditional distribution function. They also 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.

Shim *et al.* [2009] propose a new non-crossing quantile regression method using doubly penalized kernel machine that uses heteroscedastic location-scale as basic model and estimates both location and scale simultaneously by kernel functions.

Wu & Liu [2009] introduce a stepwise estimation scheme. With the current quantile regression function at a particular given level, constraints are added in the estimation procedure to ensure the next quantile regression function does not cross the current one. The procedure continues until quantile regression functions at all desired levels are obtained. One drawback of this algorithm is its dependence on the order that the quantiles are fitted.

The point of departure of Chernozhukov *et al.* [2009, 2010] is that if an original, potentially non-monotonic, estimate is available, then the rearrangement operation from variational analysis can be used to monotonize the estimate of the quantile regression curves. To this end, the authors propose monotone rearranging the original estimated curves, which are closer to the true quantile curves than the original curves in finite samples. However, the estimate of the conditional distribution function $y|\mathbf{x}$

is modified in a way which makes problematic to quantify effects of the explanatory variables.

Liu & Wu [2011] employ simple constraints on the kernel coefficients which can guarantee that the estimated conditional quantile functions never cross each other. This kernel formulation covers both linear and nonlinear models. Furthermore, the authors demonstrate that through sharing strength among different quantiles, simultaneous non-crossing quantile regressions can produce better estimation than individually estimated quantile functions.

The basic idea of Schnabel & Eilers [2013,b] is to introduce a surface on a two-dimensional domain. One axis is for the predictors, the other is for the probability p. The quantile curve for any probability is found by cutting the surface at that probability. Effectively, all possible quantile curves are estimated at the same time and the crossing problem disappears completely if the sheet is monotonically increasing with p for every variable.

Rather than directly modeling the level of each individual quantile, Schmidt [2013] begins with a single quantile (*e.g.* the median), and then add or subtract nonnegative functions (called quantile spacings) to it in order to find the other quantiles. This approach is analogous to methods for approximating intervals, where one models the midpoint and the range of the interval, rather than try to model the upper and lower bounds directly.

3. Non crossing regression quantiles

Crossings of quantile regression hyperplanes are an undesirable inconsistency that undermines the theoretical integrity of the quantile regression method and limits its usefulness in applications where monotonicity is a critical issue. We therefore attempt to force proper ordering of the quantile curves to ensure that there are no crossings over some relevant region of covariate space.

3.1. Unweighted non-crossing quantile regressions

If we apply the quantile function model (1.1) for the set of fixed k percentages $p \in P$, then we need to estimate k blocks of coefficients $\mathbf{B} = [\boldsymbol{\beta}(p_1), \boldsymbol{\beta}(p_2), \cdots, \boldsymbol{\beta}(p_k)]$. The corresponding k conditional quantile functions should verify the monotonicity requirements with respect to p.

$$sgn(p_h - p_{h-1})\, sgn\Big[\mathbf{x}^t \hat{\boldsymbol{\beta}}(p_h) - \mathbf{x}^t \hat{\boldsymbol{\beta}}(p_{h-1})\Big] \geq 0;\ h = 2, \cdots, k \quad \forall\, \mathbf{x} \in S^m \tag{3.1}$$

In the absence of further restrictions, the estimators in $\mathbf{B}$ would be obtained by solving the minimization problem (1.3) for each $p \in P$. As we have said in the previous section, crossings should never happen in theory because of the properties

of the quantile regression estimators. The question remains however how to deal with overlapping hyperplanes when such cases do occur.

Bondell *et al.* [2010] study a simple constrained version of quantile regression in which, to alleviate the crossing issue. Let $\mathbf{L} = (L_1, L_2, \cdots, L_m)$ and $\mathbf{U} = (U_1, U_2, \cdots, U_m)$ be, respectively, the vector of minimum and the vector of maximum elements observed for each explanatory variables (with the exclusion of the first columns consisting entirely of ones). To simplify the evaluation of constraints (3.1), we can transform the variables so that they range into the interval $[0, 1]$

$$v_{i,j} = \begin{cases} \dfrac{x_{i,j} - L_j}{U_j - L_j} & \text{for } j = 2, \cdots, m \\ 1 & \text{for } j = 1 \end{cases} \quad ; \; i = 1, 2, \cdots, n. \tag{3.2}$$

The domain of interest is now reduced from S^m to $D^m = \{\mathbf{v} | \mathbf{v} \in [0, 1]^m\}$. Hence a quantile regression estimate of the unknown parameters can be given by

$$\mathbf{Q}\left(\tilde{\boldsymbol{\beta}}, \mathbf{y}, \mathbf{v}\right) = \min_{\tilde{\boldsymbol{\beta}} \in D^m} \left\{ p \sum_{i | y_i \geq \mathbf{v}_i^t \tilde{\boldsymbol{\beta}}} \left(y_i - \mathbf{v}_i^t \tilde{\boldsymbol{\beta}} \right) + (1-p) \sum_{i | y_i < \mathbf{v}_i^t \tilde{\boldsymbol{\beta}}} \left(\mathbf{v}_i^t \tilde{\boldsymbol{\beta}} - y_i \right) \right\} \tag{3.3}$$

where $\mathbf{v}_i = (v_{i,1}, \cdots, v_{i,m})$. The linear programming theory assures that feasible solutions to (3.3) occur at the vertices of D^m, *i.e.* set of integers $M \subset (1, \cdots, n)$ corresponding to observations for which: $\tilde{\boldsymbol{\beta}}(p) = [\mathbf{V}(M)]^{-1} \mathbf{y}(M)$, where $\mathbf{V}$ is the matrix whose rows are $\mathbf{v}_1, \cdots, \mathbf{v}_n$. For simplicity of manipulation, it is convenient to redefine the k solution vectors of (3.3) for $p \in P$ as follows

$$\boldsymbol{\gamma}(p_1) = \tilde{\boldsymbol{\beta}}(p_1) \, ; \qquad \boldsymbol{\gamma}(p_j) = \tilde{\boldsymbol{\beta}}(p_j) - \tilde{\boldsymbol{\beta}}(p_{j-1}) \quad j = 2, \cdots, k. \tag{3.4}$$

The restrictions described in (3.1) are now equivalent to

$$\mathbf{v}_i^t \boldsymbol{\gamma}(p_h) \geq 0 \quad \forall \mathbf{v} \in D \quad h = 2, \cdots, k. \tag{3.5}$$

This condition, according to Bondell *et al.* [2010], is both necessary and sufficient to prevent overlapping hyperplanes. The linear transformation (3.2) can easily be inverted after the estimation, while retaining the properties of the quantile regression estimators. In fact, if $\tilde{\boldsymbol{\beta}}(p)$ is the vector of estimated parameters associated with $p \in P$ under (3.2), then

$$\hat{\beta}_j(p) = \begin{cases} \dfrac{\tilde{\beta}_j(p)_j}{U_j - L_j} & \text{for } j = 2, \cdots, m; \\ \tilde{\beta}_1(p) - \sum_{j=2}^{m} \left[\dfrac{\tilde{\beta}_j(p) * L_j}{U_j - L_j} \right] & \text{otherwise} \end{cases} . \tag{3.6}$$

The merit of the approach outlined by Bondell *et al.* [2010] is that the question of quantile crossings is now reduced to a linear programming problem, which can be solved via standard software.

3.2. Weighted non-crossing quantile regressions

One unrealistic assumption underlying the quantile regression model is that each point of the p-th regression quantile hyperplane provides equally reliable and valid information about the deterministic part of the response variable. We argue that quantile regression crossings are due, at least in part, to the fact that all observations are considered on the same footing although the data might not justify this. Furthermore, we claim that the use of residuals from quantile regression can be of help to avoid such shortcomings.

Consistent with this premise, we believe that a way to avoid intersections between estimated hyperplanes (over the domain of interest) is to put more emphasis on observations which are more coherent with the model (1.1) and give less importance to observations thought to be cause of irregularities. Therefore, to deal with the crossing issue, we propose to estimate the quantile regressions under the non-crossing restrictions (3.1) by adjusting fit to the following objective function

$$\mathbf{Q}(\boldsymbol{\beta}^*,\mathbf{y},\mathbf{v}) = \min_{\boldsymbol{\beta}^* \in D^m} \left\{ p_h \sum_{i|y_i \geq \mathbf{v}_i^t \boldsymbol{\beta}^*} w_{i,n} \left(y_i - \mathbf{v}_i^t \boldsymbol{\beta}^* \right) + (1-p_h) \sum_{i|y_i < \mathbf{v}_i^t \boldsymbol{\beta}^*} w_{i,n} \left(\mathbf{v}_i^t \boldsymbol{\beta}^* - y_i \right) \right\} \tag{3.7}$$

where the weighs verifies the conditions.

$$w_{i,n} > 0, \qquad \lim_{\mathbf{Q}(\tilde{\boldsymbol{\beta}},\mathbf{y},\mathbf{v}) \to 0} w_{i,n} = \omega > 0 \quad \text{for} \quad i = 1, \cdots, n. \tag{3.8}$$

The magnitude of $w_{i,n}$ quantifies the suitability of the information contained in the i-th observation relatively to the k regression hyperplanes fitting the n data points. Strictly positive weights are strongly recommended by Koenker [2013][p. 17] since a null weight is ambiguous. Moreover, the weights should tend to be equal when the fit tend to be ideal. Note that the weighted version of the objective function (3.7) can be solved by applying the unweighted algorithm to the responses and explanatory variables defined by $r_{i,h} = w_{i,n}$ and $y_i, \mathbf{z}_i = w_{i,n}\mathbf{v}_i$ for $i = 1, \cdots, n$. Therefore, problem (3.7) can be reformulated as follows

$$\mathbf{Q}\left(\tilde{\boldsymbol{\beta}}, \mathbf{r}, \mathbf{z}\right) = \min_{\tilde{\boldsymbol{\beta}} \in R^m} \left\{ p_h \sum_{i|r_i \geq \mathbf{z}_i^t \tilde{\boldsymbol{\beta}}} \left(r_i - \mathbf{z}_i^t \tilde{\boldsymbol{\beta}} \right) + (1-p_h) \sum_{i|r_i < \mathbf{z}_i^t \tilde{\boldsymbol{\beta}}} \left(\mathbf{z}_i^t \tilde{\boldsymbol{\beta}} - r_i \right) \right\}. \tag{3.9}$$

This method is particularly convenient because it involves adding only the estimated weights to a computer program with a weighting option. In practice, the computation of non-crossing weighted quantile regression can be efficiently accomplished by exploiting the same software developed for Bondell *et al.* [2010]. Our approach presupposes that the weights are fixed and known in advance. For example, they can hold information about the reliability of imputed values or values derived from previous experience or from source known to be polluted by errors of measurement.

Also, weights can be used to take into account of cases with the same values on all variables. In practice, however, the assumption of known weights rarely holds so estimated weights must be used instead. There are many ways to estimate $\mathbf{w}$. We base our choice on the idea that the weight for each observation should be inversely related to the size of the corresponding disturbance $\hat{e}_{i,h} = y_i - \hat{y}_{i,h}$ where $y_i, h = 1, \cdots, k$ is the i-th value of the response in a sample of n points and $\hat{y}_{i,h}$ is some estimate of $y_i|\mathbf{x}_i$ for the h-th quantile.

Let $\hat{\mathbf{E}} = (\hat{\mathbf{e}}_1, \cdots, \hat{\mathbf{e}}_k)$ the $(n \times k)$ matrix with columns given by the n estimated residuals $\hat{e}_{i,h}, i = 1, \cdots, n; h = 1, \cdots, k$ of the non-crossing quantile regression associated with the k estimated hyperplanes. This implies that the unweighted non-crossing estimators of Bondell *et al.* [2010] are taken as a benchmark against which to compare weighted non-crossing quantile regressions. Furthermore, let $\bar{\mathbf{e}} = (\bar{\mathbf{e}}_1, \cdots, \bar{\mathbf{e}}_k)$ represent the $(k \times 1)$ vector of averages and $\hat{\mathbf{\Sigma}}_e$ the $(k \times k)$ matrix of variance-covariances of the k columns of $\hat{\mathbf{E}}$. It must be observed that $\bar{\mathbf{e}}_h = n^{-1} \sum_{i=1}^{n} \hat{e}_{i,h}, h = 1, \cdots, k$ is, in general, different from zero, unless the distribution of conditional residuals is symmetrical.

In order to quantify the potential impact on parameter estimation of the i-th observation $(y, \mathbf{x}_i)$ we use the Mahalanobis distance

$$d_i = \sqrt{(\dot{\mathbf{e}}_i - \bar{\mathbf{e}})^t \hat{\mathbf{\Sigma}}_e^{-1} (\dot{\mathbf{e}}_i - \bar{\mathbf{e}})} \quad i = 1, \cdots, n \tag{3.10}$$

where $\dot{\mathbf{e}}_i$ is the i-th row of $\hat{\mathbf{E}}$ and $\hat{\sigma}_{h,h'} = (n-1)^{-1} \sum_{i=1}^{n} (\dot{e}_{i,h} - \bar{\mathbf{e}}_h)(\dot{e}_{i,h'} - \bar{\mathbf{e}}_{h'})$ for $h, h' = 1, \cdots, k$. We note that if a singular $\hat{\mathbf{\Sigma}}_e$ is encountered, (3.10) can be modified by using a generalized inverse to obtain the weights (see Ben-Israel & Greville [2003]).

There is a range of possibilities for converting distances into weights. An exponential transformation is especially appealing to us because of its simplicity.

$$w_i = \exp\{-\tau d_i\} \qquad i = 1, \cdots, n \tag{3.11}$$

where $\tau \geq 0$ is a tuning parameter that may be varied to modify the influence of the distances. Increasing values of τ make observations which are at distance one from the vector of averages $\bar{\mathbf{e}}$ progressively less relevant. For a given $\tau > 0$, weights decrease as distances from $\bar{\mathbf{e}}$ increase. From another point of view, noting the resemblance between (3.11) and the density function of an exponential random variables, τ can be thought to be similar to the inverse of the expected uncertainty contained in a sample of random distances. Constant τ can be chosen arbitrarily in principle. Based on empirical experience with real as well as simulated data we suggest applying the *optimize* function offered in Base-R (see R Core Team, 2013). The method used is a combination of golden section search and successive parabolic interpolation that searches a specified interval from lower to upper for a minimum. A solution to (3.9) will be considered feasible if and only if $\mathbf{Q}(\boldsymbol{\beta}^*, \mathbf{r}, \mathbf{z}) \leq \mathbf{Q}\left(\tilde{\boldsymbol{\beta}}, \mathbf{r}, \mathbf{z}\right)$.

For what concerns large sample properties of weighted non-crossing quantile regressions, consider a set of percentages $p_1 < p_2 < \cdots, < p_k$ such that $p_h \in [\epsilon, 1-\epsilon]$ for $h = 1, \cdots, k$ and $0 < \epsilon < 0.5$ and assume:

1. The matrix $n^{-1}\mathbf{X}^t\mathbf{X}$ is positive definite.
2. The conditional densities $f_{y_i|\mathbf{x}}$ are differentiable with respect to y_i for every $\mathbf{x}$ and each $i = 1, \cdots, n$.
3. For $0 < \epsilon < 1$, there exist constants $a > 0;\ b, c < \infty$ such that

$$a \leq f_{y_i|\mathbf{x}}\left[Q_{y_i|\mathbf{x}}(p)\right] \leq b; \qquad \left|f'_{y_i|\mathbf{x}}\left[Q_{y_i|\mathbf{x}}(p)\right]\right| \leq c$$

uniformly for $\mathbf{x} \in D^m$, $\epsilon \leq p \leq (1-\epsilon)$ and uniformly in $i = 1, \cdots, n$.

Under the above conditions, Bondell *et al.* [2010] prove that the estimator obtained via (3.7) is asymptotically equivalent to the unconstrained quantile regression estimator, regardless of the choice of a weighting systems $w_{i,h}, i = 1, \cdots, n;\ h = 1, \cdots, k$. Furthermore, in another theorem, the authors show that inference for the $\sqrt{n}$-consistent constrained quantile regression can be achieved by using the known asymptotic results for classical quantile regression.

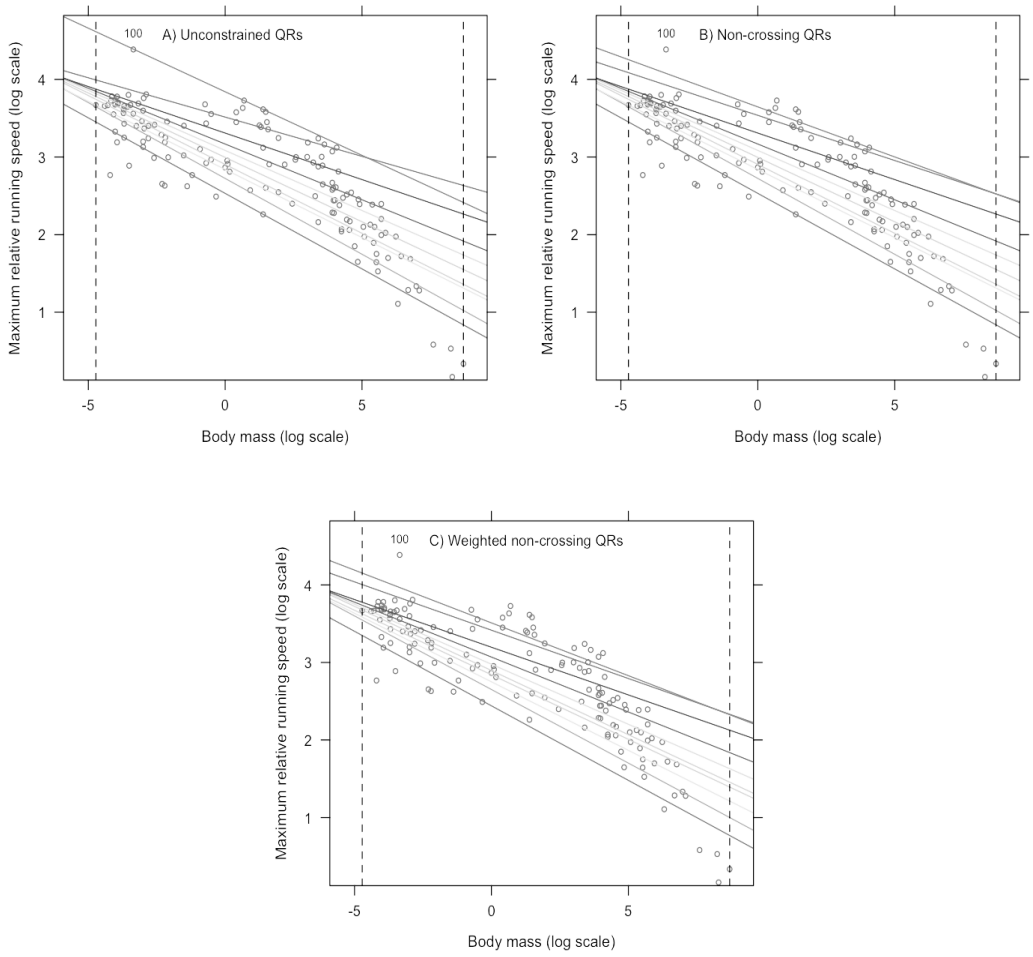
4. Experimental results

The experiments presented here look for evidence that incorporation of a weighting systems into the core of the non-crossing quantile regression procedure can lead to an alternative and (at least on specific occasions) better mechanism for fitting multivariate data. In this section, we use three examples to compare three different algorithms: unconstrained, unweighted non-crossing (UNC), weighted non-crossing (WNC) for the quantile regression, and thereby show the advantage of our new method of quantile regression.

Our first example is based on data from Iriarte-Díaz [2002]. The author discusses the relationship between maximum relative running speed (body length/second) and body mass (kg) concerning $n = 142$ species of terrestrial mammals, in order to evaluate whether the relative locomotor performance shows a differential scaling depending on the range of mass analyzed. Overall, maximum relative running speed decreases with increasing body mass. Figure 1 illustrates the results of application of the three different techniques considered in the present paper for $p \in P\,(0.50 : 0.95, by 0.05)$.

From graph B, it is apparent that the computation method proposed by Bondell *et al.* [2010] avoids the intersections which are present in graph A, at least within the domain delimited by the vertical dotted lines traced at $\min x$ and $\max x$. Our method (graph C) generates regression lines that not only bypass crossings, but also gather near the center of the observed data points. It must be noted, in fact, that

FIG 1. *Relationship between body mass and maximal relative running speed.*



there is an entity which does not match the general impression: 100 corresponding to heteromyid rodent (*Dipodomys merriami*). The bias attributable to this outlier can be noticed looking at the highest two lines in graph A and B. In the former, there is an intersection clearly due to the carry-over effect from the isolated point. In the latter, the problem of crossing is solved, but some of the lines remain close to the outlier. In graph C the influence of the outlier has been removed. The accumulation of lines around the center is presumably due the fact that the relationship between maximum running speed and body mass is curvilinear rather than linear.

To assess the difference in efficacy between different methods of estimation, we evaluate the behavior of the global absolute errors affecting the various regression methods. In particular, Table 1 compares the mean, the maximum and the minimum sum of absolute errors associated with the $k = 10$ quantile regression hyperplanes. The findings in Table 1 reveal that weighted non-crossing quantile regressions attains a better performance than the standard procedure with respect the absolute

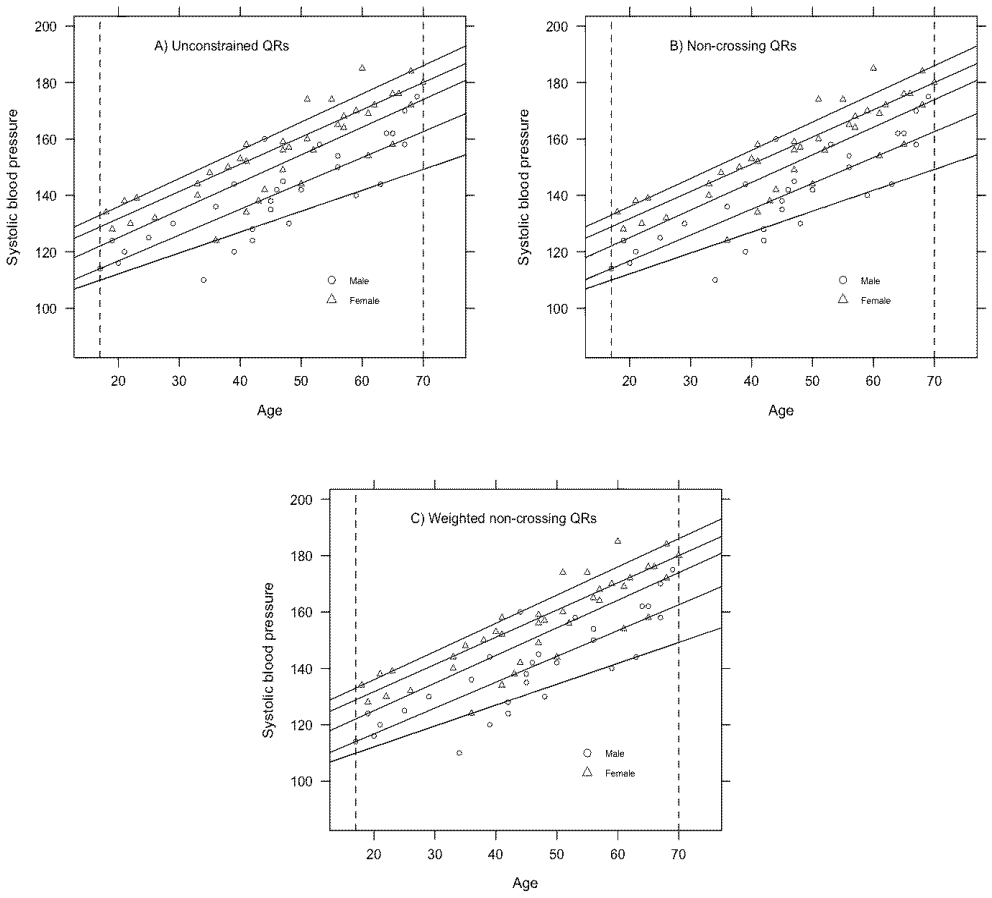
TABLE 1
Fitting results of various estimation methods

	Unconstrained QRs	Non-crossing QRs	Weighted Nc QRs
Mean absolute error	0.5055	0.5076	0.3961
Max absolute error	0.8103	0.8219	0.5479
Min absolute error	0.3417	0.3417	0.3547

residuals. The unweighted non-crossing technique does not improve, from a fitting point of view, upon unconstrained quantile regressions.

As the second example, we analyze the data set *sbp* included in the package *multcomp* of *R* for the percentages $(0.10, 0.25, 0.50, 0.75, 0.90)$. The data set refers to systolic blood pressure (in mmHg), age (in years) and gender of $n = 69$ people. In Figure 2 it is shown that, in absence of outliers in the data and non-crossing lines, the three estimation methods behave similarly.

FIG 2. *Relationship between age and systolic blood pressure.*



The well-known *Housing Data Set* which is available online at http://lib.stat.cmu.edu/datasets/boston_corrected.txt is considered for the third example. The data comprises $n = 506$ observations for 13 predictor variables, and one response variable, corrected median value of owner-occupied homes (CMEDV). The data set was analyzed by Harrison & Rubinfeld [1978] who wanted to find out whether "clean air" had an influence on house prices. For simplicity, we excluded the categorical variable RAD and the Charles River dummy variable (because there are too few on one status) and considered $m = 11$ predictor variables.

We select virtual random samples without repetition of $n \in (120, 240, 360, 480)$ observations from the total data set. The results are reported in Table 2 where each entry is an average across $L = 100$ experiments of the same type.

TABLE 2
Fitting results of various estimation methods.

	Unconstrained QRs			Non-crossing QRs			Weighted Nc QRs		
n	mean	max	min	mean	max	min	mean	max	min
120	4.7013	8.5361	2.9504	4.7150	8.4104	3.0048	3.6294	4.1298	3.2905
240	4.8995	9.0573	3.0772	4.9320	9.1909	3.1003	3.7013	4.2762	3.3648
360	4.9112	9.2036	3.0871	4.9100	9.2307	3.0989	3.7188	4.2760	3.3603
480	4.9537	9.3647	3.1062	4.9786	9.5880	3.1138	3.7984	4.5642	3.3409

The weighted non-crossing quantile regressions yield average absolute errors systemically better than those of the other methods. It appears that, the adjustments caused by the unweighted restrictions on the intersection of hyperplanes of the ordinary estimates have resulted in relatively minor modifications to the extremes quantile regressions. The adjustments are more substantial for weighted non-crossing regressions, and these seem to be concentrated in the central and higher percentages where the most pronounced reduction of residual reductions is observed.

The quality of the fitting expressed by the columns of Table 2 does not improve with increased sample size. Rather it seems getting worse although at no time is it sharp. This could be explained by the clear clustered structure of the observations.

5. Discussion and Conclusions

Conditional quantile functions offer simple and flexible models for the stochastic component of a regression and enable us to obtain reasonable estimates in the presence of a broad range of departures from Gaussianity. (See in particular Parzen, 1979 and Gilchrist, 2006). However, the interpretability of QR estimates deteriorates when conditional quantile functions cross or overlap.

Our aim in this paper is to introduce a new methods of estimation for the parameters of quantile regressions that avoids the problem of crossing quantile curves.

Based upon the work Bondell *et al.* [2010], a weight is attached to each observation inversely related to the estimated disturbances associated with the unweighted quantile regressions. We are convinced that the influence of disturbances corresponding to a given observation decreases exponentially with the Mahalanobis distance from their centroid. This scheme can be particularly effective when the intersection of hyperplanes is most probably due to the presence of outlying entities.

The estimation of multiple non-crossing quantile regressions is enforced by requiring nothing more than lower quantile levels do not cross higher quantile levels. This gives rise to a set of inequalities that should be all satisfied. Inequalities can be considered *a priori* pieces of information about the true parameters that restrict the original parameter space. It is known that, under general conditions, the estimate $\hat{\boldsymbol{\beta}}(p_h)$ has optimal properties for the h-th conditional distribution and this is also true for any $h = 1, \cdots, k$. Since the unconstrained estimation procedure does not use the fact that $\boldsymbol{\beta}(p_h), h = 1, \cdots, k$ lie in the reduced space, one might wonder if using such conditions gives a gain in efficiency. This is not necessarily so (see Rothenberg [1973][p. 55-57] for the case of the linear least squares estimators). In this respect, Takeuchi *et al.* [2006] note that, after enforcing the constraints, the quantile property may not be guaranteed. This is because we try to optimize both for the quantile property and the non-crossing property (in relation to other quantiles). Hence, the final outcome may not empirically satisfy the quantile property. The question then is to find the best way of satisfying the constraints without worsening the property of the regression quantile estimators.

We have shown that our method, because of the introduction of an efficient system of weights, is successful at determining quantile regression hyperplanes that do not cross in the convex hull of the explanatory variables. The results presented in this paper support this view. There are still many unknown aspects of our methodology; for example, what is the efficiency of parameter estimates for clean data (absence of outliers or Gaussian disturbances), what is the power function of the test statistics, and what is the bias in parameter estimates when data are affected by specific forms of heteroscedastic errors. These problems can be addressed through asymptotics for large samples and via a diffuse Monte Carlo simulation plan evaluation for finite samples. These will be topics for further study. Two other potential directions for future research should be considered: to devise a multistep mechanism for building more effective weights and to establish test statistics which help which help to decide on goodness of fit for systems of quantile regressions on the same data set.

References

Bassett, G. and Koenker, R. "An empirical quantile function for linear models with iid errors" *Journal of the American Statistical Association* 77, 407–415 (1982).

Bassett, G. and and Koenker, R. "Strong consistency of regression quantiles and related empirical processes" *Econometric Theory* 2, 191–201 (1986).

Ben-Israel, A. ad Greville, T. N. E. *Generalized Inverses: Theory and Applications, 2nd ed.* Springer-Verlag, New York (2003).

Bondell, H. D. and Reich, B. J. and Wang, X. "Noncrossing quantile regression curve estimation" *Biometrika* 97, 825–838 (2010). Software available at http://www4.stat.ncsu.edu/~bondell/Software/NoCross/NoCrossQuant.R

Buchinsky, M. "Recent advances in quantile regression models: a practical guideline for empirical research" *The Journal of Human Resources* 33, 88–126 (1998).

Chernozhukov, V. and Fernández-Val, I. and Galichon, A. "Improving point and interval estimators of monotone functions by rearrangement" *Biometrika* 96, 559–575 (2009).

Chernozhukov, V. and Fernández-Val, I. and Galichon, A. "Quantile and probability curves without crossing" *Econometrica* 78, 1093–1125 (2010).

Dette, H. and Volgushev, S. "Non-crossing non-parametric estimates of quantile curves" *Journal of the Royal Statistical Society, Series B* 70, 609–627 (2008)

Gilchrist, W. G., "Modeling and fitting quantile distributions and regressions", Sheffield Hallam University (2006).

Harrison, D. and Rubinfeld, D. L. "Hedonic prices and the demand for clean air" *Journal of Environmental Economics & Management* 5, 81–102 (1978).

He, X. "Quantile Curves without Crossing" *The American Statistician* 51, 186–192 (1997).

Heagerty, P. J. and Pepe, M.S. "Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children" *Applied Statistics* 48, 533–551 (1999).

Iriarte-Díaz, J. "Differential scaling of locomotor performance in small and large terrestrial mammals" *The Journal of Experimental Biology* 205, 2897–2908 (2002).

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

Koenker, R. (2012) quantreg package in R, version 4.91. (Available from http://cran.r-project.org/ web/packages/quantreg/index.html.)

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

Koenker R., and D'Orey V. "Algorithm AS 229: computing regression quantiles" *Applied Statistics* 36, 383–393 (1987).

"Remark AS R92: a remark on algorithm AS 229: computing dual regression quantiles and regression rank scores" *Applied Statistics* 43, 410–414 (1994).

Koenker, R. and Geling, O. "Reappraising medfly longevity: a quantile regression survival analysis" *Journal of the American Statistical Association* 96, 458–468 (2001).

Koenker, R. "Quantreg: quantile regression" *R package version 5.05* (2013). Available at http://CRAN.R-project.org/package=quantreg

Liu, Y. and Wu, Y. "Simultaneous multiple non-crossing quantile regression estimation using kernel constraints" *Journal of Nonparametr Statistics* 23, 415–437 (2011).

Machado, J. and Mata, J. "Counterfactual decomposition of changes in wage distributions using quantile regression" *Journal of Applied Econometrics* 20, 445–465 (2005).

Melly, B. "Decomposition of differences in distribution using quantile regression" *Labour Economics* 12, 577–590 (2005).

Neocleous, T. and Portnoy, S. "On monotonicity of regression quantile functions" *Statistics and probability letters* 78, 1226–1229 (2007).

Parzen, E. "Nonparametric statistical data modelling" *Journal of the American Statistical Association* 74,105–131 (1979).

Portnoy, S. "Asymptotic behavior of the number of regression quantile breakpoints" *SIAM Journal on Scientific and Statistical Computing* 12, 867–883 (1991).

R Core Team *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ (2013).

Rothenberg, T. J. *Efficient estimation with a priori information*. Yale University Press, New Haven (1973).

Schmidt, L. D. W. "Quantile spacings: a simple method for the joint estimation of multiple quantiles" (2013). Available http://ssrn.com/abstract=2220901

Schnabel, S. and Eilers, P. H. C. "Simultaneous estimation of quantile curves using quantile sheets" *Advances in Statistical Analysis* (2013a).

Schnabel, S. and Eilers, P. H. C. "A location-scale model for non-crossing expectile curves" *Stat* 2, 171–183 (2013b).

Schlossmacher, E.J. "An iterative technique for absolute deviations curve fitting" *Journal of the American Statistical Association* 68, 857–859 (1973).

Seneta, E. and Steiger, W. L. "A new LAD curve-fitting algorithm: slightly overdetermined equation systems in L_1" *Discrete Applied Mathematics* 7, 79–91 (1984).

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

Takeuchi, I. and Furuhashi, T. "Non-crossing quantile regressions by SVM" *Proceedings. 2004 IEEE International Joint Conference on Neural Networks* 1, 401–406 (2004). Budapest, Hungary, 25-29 July, 2004

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

Tokdar, S. T. and Kadane, J. B. "Simultaneous linear quantile regression: a semiparametric Bayesian approach" *Bayesian Analysis* 7, 51–72 (2012).

Wu, Y. and Liu, Y. "Stepwise multiple quantile regression estimation using non-crossing constraints" *Statistics and Its Interface* 2, 299–310 (2009).

Yu, K. and Jones, M. C. "Local linear quantile regression" *Journal of the American Statistical Association* 93, 228–237 (1998).

Zhao, Q. "Restricted regression quantiles" *Journal of Multivariate Analysis* 72, 78–99 (2000).