Original Research Paper

Quantile Regression Estimation Using Non-Crossing Constraints

Ilaria Lucrezia Amerise

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

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 a set of explanatory variables. More specifically, we investigate the advantage of using a new method to estimate a bunch of non-crossing quantile regressions hyperplanes. The main tool is a weighting system of the data elements that aims to reduce the effect of contamination of the sampled population on the estimated parameters by diminishing the effect of outliers. The performances of the new estimators are 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 imputed data) the method of weighted non-crossing quantiles will lead to estimators with good robustness properties.

Keywords: Conditional Quantiles, Monotonicity Problem, Estimation Under Constraints

Introduction

Methodology and Estimation

A typical investigation in statistical analysis consists of the linear regression of one response variable on one or more predictor variables, all of which are observed on a sample of entities. The rational is that by establishing a linear relationship between them, knowledge of the value of predictor variable 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 us to examine and compare different levels of response given the variation in the independent variables by considering simultaneously the conditional quantile functions 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 covariates $\mathbf{x}$. In short, y is defined as the smallest real value such that the probability of obtaining smaller values of Y is at least p. The quantiles are the values that divide the total probability into parts. Values of interest are themedian 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 = 0.25 h, h = 1,2,3), deciles (p = 0.1 h, h = 1,···, 9) and so on. Quantiles could be considered also at irregularly spaced over the (0,1) interval of probabilities. Conditional quantile functions offer simple and flexible models for the stochastic component of a regression and enables us to obtain reasonable estimates in the presence of a broad range of departures from Gaussianity (Parzen, 1979; Gilchrist, 2006).

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

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

with $x_i = (\mathbf{x}_{i,1}, x_{i,2}, \cdots, x_{im})$, $i = 1,2,..., n$, being a sequence of $m \times 1$ vectors of known covariates and $n > m$. For each p, the magnitude and the sign of the effect of a given regressor can be compared with the effect at the other quantiles. The vector $\beta(p) \in R^m$ contains p-specific coefficients whose estimate should be obtained from sample data. The $\hat{\beta}_j (p), j = 1,2,\cdots, m$ can be interpreted as the trade-off $\dfrac{\partial Q_p(Y|\mathbf{x})}{\partial x_j}$ 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 covariate, when all of the other covariates are held constant. If $\beta(p)$ is a constant β for each p, the model reduces to the standard conditional expectation model $y_i = \mathbf{x}_i^t \beta$ with constant variance errors $\sigma(e)$.

Denote $X = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ the $n \times m$ known design matrix with columns $\mathbf{x}_i$, $i = 1, 2, \cdots, m$. We assume that the covariates include an intercept term; therefore, the first column of $\mathbf{X}$ consist entirely of ones and that $\mathbf{X}$ has rank m. In addition, let $S^m \subset R^m$ the compact domain of the covariates over which the model holds; this means that, for each experimental condition x, an observation y is available according to the model (1). We assume further that y and $\mathbf{X}$ are observed with no error and that different observations are independent. Moreover, we assume that $e_1, e_2, \cdots, e_n$ are independent random errors with quantile function $Q_p(e)$. The quantile function is left unspecified; we only require the verification of the following constraint:

$$Q_p\left(e_{p,i} \mid \mathbf{x}_i\right) = 0 \tag{2}$$

which implies that the conditional p-th quantile of e_i is null for each i. It follows that the p-th conditional quantile of $y \mid x_i$ is given by:

$$Q_p\left(y_i \mid \mathbf{x}_i\right) = Q_p\left[\mathbf{x}_i^t \beta(p) + e_{p,i} \mid \mathbf{x}_i\right]$$
$$= Q_p\left[\mathbf{x}_i^t \beta(p) \mid \mathbf{x}_i\right] + Q_p\left[e_{p,i} \mid \mathbf{x}_i\right] = \mathbf{x}_i^t \beta(p). \tag{3}$$

Not that $Q_p\left[\mathbf{x}_i^t \beta(p) \mid \mathbf{x}_i\right] = \mathbf{x}_i^t Q_p\left[\beta(p) \mid \mathbf{x}_i\right] = \mathbf{x}_i^t \beta(p)$ by the definition of the conditional expectation. It is worth noting that there is no assumption on identical distributions and that model (1) allows the errors to change as a function of $\mathbf{X}$ and, thus, various form of and local noise rates can be accomodated.

The first thing you need to understand is that there are two classes of optimization problems: Function minimization and mathematical programming. Both seek an optimum and both involve constraints, but mathematical programming begins with the constraints. It emphasizes solving the constraints and then looks for the best solution.

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

$$\mathbf{Q}(\beta, \mathbf{y}, \mathbf{x})$$
$$= \min_{\beta \in R^m}\left\{ p \sum_{i \mid y_i \geq \mathbf{x}_i^t \beta} \left(y_i - \mathbf{x}_i^t \beta\right) + (1-p) \sum_{i \mid y_i < \mathbf{x}_i^t \beta} \left(\mathbf{x}_i^t \beta y_i\right) \right\}. \tag{4}$$

The minimizing $\hat{\beta}$ determine a m dimensional hyperplane given by $\mathbf{x}_i^t \hat{\beta}(p)$ that best fits the n observations. All points 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 and 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 which, in turn, can be reformulated as the minimization of a linear function subject to linear constraints:

$$\min_{\beta(p) \in R^m; \mathbf{r}, \mathbf{s} \in R^n} \left[p\mathbf{r}^t \mathbf{u}_n \, (1-p)\mathbf{s}^t u_n \, \mathbf{0}_n^t \right] \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \\ \beta(p) \end{bmatrix} \tag{5}$$

subject to:

$$\begin{bmatrix} -\mathbf{I}_{n,n} & \mathbf{I}_{n,n} & -\mathbf{X} \\ \mathbf{I}_{n,n} & -\mathbf{I}_{n,n} & \mathbf{X} \\ \mathbf{I}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{I}_n & \mathbf{0}_{n,n} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \\ \beta(p) \end{bmatrix} \geq \begin{bmatrix} \mathbf{y} \\ \mathbf{y} \\ \mathbf{0}_n \\ \mathbf{0}_n \end{bmatrix} \tag{6}$$

where:
$\mathbf{u}_n$ = The $(n \times 1)$ vector of ones
$\mathbf{I}_{n,n}$ = The identity matrix of order n
$\mathbf{0}_{n,n}$ = The $(n \times n)$ matrix of zeros
$\mathbf{0}_n$ = The $(n \times 1)$ vector of zeros

The linearity of the objective function and constraints implies that the solution has to lie in one of the vertices of the polyhedron defined in (5). The advent of modern linear programming techniques in the later 1940s lead to fast and efficient algorithm. In fact, it may be the computational complexity of least absolute regression, as well as the analytical intractability in a statistical setting, that forced this approach to take the back seat to least squares in multiple linear regression Seneta and Steiger (1984). Here we will not go deep into the computational aspects of quantile regression, but just recall that the availability of efficient linear programming algorithms (Koenker and D'Orey, 1987; 1994). For large data sets the interior point algorithm written by Portnoy (1991) is recommended.

Buchinsky (1998) notes that the $m \times 1$ vector of first-order conditions for solving the problem in (4) is given by:

$$Un(\beta, p) = n^{-1} \sum_{i=1}^{n} \left[p - \frac{1}{2} + \frac{1}{2} sgn\left(y_i - \mathbf{x}_i^t \beta\right) \right] \mathbf{x}_i \approx 0 \tag{7}$$

where, the *sgn*(.) (*signum*) function takes the values $-1,0,+1$ according to whether its argument is negative, zero or positive. The approximation symbol in (7) emphasizes the fact that since $U_n(\beta, p)$ is discontinuous function of $\beta(p)$, it may not have an exact solution. However, for $n \rightarrow \infty$ the expression on the left hand side of (7) converges to zero.

Let $M = (r_1, r_2, \cdots, r_m)$ be a subset of m distinct integers from $1, 2, \cdots, n$ determining a combination of rows of the design matrix such that $\mathbf{X}(M)$ has rank m. According to Koenker and Bassett (1978) (Theorem 3.1) the solutions to (4) have the form:

$$
\begin{aligned}
&\left[\mathbf{X}(M)\right]^{-1}\mathbf{y}(M) \\
&\mathbf{r} = \max\left\{\mathbf{y} - \mathbf{X}\beta(p), \mathbf{0}_n\right\}, \quad \mathbf{r}_M = \mathbf{0}_M \\
&\mathbf{s} = -\min\left\{\mathbf{y} - \mathbf{X}\beta(p), \mathbf{0}_n\right\}, \quad \mathbf{s}_M = \mathbf{0}_M
\end{aligned}
\tag{8}
$$

where, $\mathbf{0}_n$ and $\mathbf{0}_M$ are vectors of n and M zeros, respectively. B is not empty. Moreover, if the quantile function of the error term is continuous, then the system (4) has a unique solution $\hat{\beta}(p)$ provided that:

$$
\mathbf{u}_k^t < \sum_{i \in r}\left\{0.5\left[1 - sgn\left(yi - \mathbf{x}_i^t\hat{\beta}(p)\right)\right] - p\right\}\mathbf{x}_i\mathbf{X}(M)^{-1} < p\mathbf{u}_k^t.
\tag{9}
$$

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

$$
\frac{n^-}{n} \le p \le 1 - \frac{n^+}{n}
\tag{10}
$$

where n^-, n^+ indicate, respectively, the number of negative and positive residuals. If the solution of (4) is unique then all inequalities are strict. Furthermore, if the quantile function of the error terms is continuous then there are exactly m residuals with value zero. To clarify this point, consider the formulation of (1) with $m = 1$ (intercept-only model). The resulting value of $\hat{\beta}(p)$ (a scalar in this case) that minimizes (4) occurs only when $\hat{\beta}(p)$ is equal to the p-th quantile of the y. 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 $n \rightarrow \infty$, we could estimate an increasing number of quantile regressions; in practice, there may be at most $3n$ distinct regression solutions for p unequally spaced on the interval $[0,1]$ (Koenker and D'Orey, 1987). 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 increases. In finite samples, Portnoy (1991) shows that the number of distinct quantile regressions is $O(nlogn)$. This opportunity is particularly useful when the regressors 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 focuses on how the covariates affect both the tails and the central parts of the conditional distribution. There will have to be a thorough understanding of the research situations in which quantile regression will be used in order both to understand its own limitations and to exploit its full potential.

Inference and asymptotic theory of quantile regression is not discussed in this paper because of the availability of a vast literature on the subject. The survey in Koenker (2005)[Ch.3 and Ch. 4] is particularly effective in this sense. We shall therefore confine ourselves to highlighting a few issues about de estimated error terms.

Crossing Quantile Regressions

Quantile regression estimates are robust in presence of observations that are far in the direction of the response variable (note that this does not apply to outliers of covariates). This is an attractive property, at least in part *attributable* to "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 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 gap between percentages p 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 $0 < p_1 < p_2 < \cdots < p_k < 1$ and the estimation of the parameters separately for each quantile. However, when several conditional quantiles are treated, it is not unusual that the estimated parameters generate non parallel hyperplanes.

Quantile regression hyperplanes in R^m are defined by the real coefficients $\beta_j(p), j = 1, \cdots, m$ where the set:

$$\left\{ \mathbf{x} \in R^m \left| \sum_{j=1}^{m} \beta_j(p) x_j - y = 0 \right. \right\} \quad with \quad \beta_j(p) \neq \mathbf{0}. \tag{11}$$

Note that two hyperplanes equations in R^m form the same hyperplane if and only if they differ by a multiplicative factor not equal to zero. In general, a violation of the monotonicity condition occurs if for two quantiles p_1, p_2 and a ($m \times 1$) vector $\mathbf{x}$ we have:

$$\mathbf{x}^t \beta(p_1) = \mathbf{x}^t \beta(p_2) \quad for \, p_1 \neq p_2, \quad \mathbf{x} \in R^m. \tag{12}$$

If the vectors $\beta(p_1)$ and $\beta(p_2)$ are linearly independent, then we have 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 $x_{r,s}$ 6 = r. At this point there are ($m-2$) free unknowns. Any two non parallel hyperplanes intersect in one hyperplane of dimension ($n-2$). When k, the number of quantiles, increases and a limited amount of data is available, the phenomenon of crossing becomes much more likely.

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

$$Q_p(y_i | \mathbf{x}_i) = \beta_0(p) + \beta_1(p) x_i \quad for \, i = 1, 2, \cdots, n. \tag{13}$$

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. This simply implies that $y|x_i$ 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 covariate effects). In this sense, Koenker and Geling (2001) suggest introducing additional covariates to avoid crossing. For example, we can vary the specification of the model for each quantile by adding and subtracting a positive covariate:

$$\begin{aligned} Q_p(y_i | \mathbf{x}_i) &= \beta_0(p) + \beta_1(p) x_i \\ &\pm \exp[\beta_2(p) x_i] \quad for \, i = 1, 2, \cdots, n. \end{aligned} \tag{14}$$

This expression incorporates two quantile regressions which never cross one another and do not cross (13). 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 intercept 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 he covariates is R^m. Tokdar and Kadane (2012) build a model of quantile regression monotonically increasing in $p \in [0,1]$ obtained by reparametrizing the elements of $\beta(p)$ as linear combinations of two monotonically increasing curves.

Bassett and 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 $Q_p = (y_i | \bar{\mathbf{x}}) = \bar{\mathbf{x}}^t \beta$, which is a monotone jump function of p on the interval [0,1]. Moreover, $Q_p(y_i | \mathbf{x})$ must be monotonic in p in a neighborhood of $\mathbf{x} = \bar{\mathbf{x}}$. Thus, incidence of crossing generally occur only in outlying regions of the observed covariate space S^m. On the other hand, we should ignore points close to the boundary or lying outside S^m unless the data set include sufficient observation the extreme regions of the design space to allow a reliable computation of quantiles. Schnabel and Eilers (2013a) 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 and Mata (2005) recall the theoretical results of Bassett and Koenker (1982) (Theorem 3.2) and Bassett and 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 design matrix $\mathbf{X}$ and the estimated hyperplanes are not guaranteed to be parallel. 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.

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 which are capable of fitting the data appropriately and several attempts at this have been made since the late 1990s. The literature 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 power transformation to restrict regression quantiles (Heagerty and Pepe (1999). The second imposes certain restrictions on the space of possible solutions to conditional quantiles. The restricted regression quantile curves are easy to compute, but do not suffer from the problem of quantile reversal in certain areas of the design space.

Yu and 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 quantile. 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 and 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 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 $y_i = \mathbf{x}_i^t \beta(p)$ is estimated along a grid of k different p-values whose mesh is sufficiently dense (a mesh size of order $O\left(n^{-(0.5+\varepsilon)}\right)$ will work). In the second stage, quantiles of the sample of $k * n$ estimates $\mathbf{x}_i^t \hat{\beta}(p_h), h = 1, \cdots, k; i = 1, \ldots, n$ are computed by weighting each element by $(p_h - p_{h-1})$. The result is the estimation of the unconditional quantiles of y.

Neocleous and 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 is asymptotically equivalent to the usual regression quantile estimator. Dette and 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 which uses heteroscedastic location-scale model as basic model and estimates both location and scale simultaneously by kernel function. Wu and 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 till 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 original estimate of the quantile regression curve. In this sense, the authors propose monotone rearranging the original estimated curves, which is closer to the true quantile curve than the original curve 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 covariates.

Bondell *et al*. (2010) study a simple constrained version of quantile regression in which, to alleviate the crossing issue, it is proposed to give a different positive weight to each $\mathbf{x}_i^t \beta(p)$. The weight is unique to the p-th hyperplane but it is common to all the n observations.

Liu and Wu (2011) employ simple constraints on the kernel coefficients which can guarantee the estimated conditional quantile functions never cross each other. This kernel formulation covers both linear and nonlinear models. Furthermore, the authors demonstrate how that through sharing strength among different quantiles, simultaneous noncrossing quantile regression can produce better estimation than individually estimated quantile functions.

The basic idea of Schnabel and Eilers (2013a; 2013b) is to introduce a surface on a twodimensional domain. One axis is for the covariates, 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 covariate value.

Rather than directly modeling the level of each individual quantile, begins with a single quantile (usually the median) and then add or subtract nonnegative functions (called quantile spacings) to it in order to find the other quantiles. His 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.

Weighted Quantile Regressions

Crossings of quantile regression hyperplanes is an undesirable inconsistency which undermines the theoretical integrity of the quantile regression method and limits its usefulness in applications where monotonicity is

a critical issue, such as prediction intervals for forecast. We therefore attempt to force proper ordering of the quantile curves to ensure that there is no crossings over some relevant region of covariate space.

Non-Crossing Quantile Regressions

If we apply the quantile function model (1) for $p \in P = (0 < p_1 < \cdots, p_k < 1)$ then we need to estimate k sets of coefficients $\mathbf{B} = [\beta(p_1), \beta(p_2), \cdots, \beta(p_k)]$. The corresponding k conditional quantile functions should verify, in a natural way, the monotonicity require-ments with respect to the percentage p:

$$sgn(p_h - p_{h-1}) sgn\left[\mathbf{x}^t \hat{\beta}(p_h) - \mathbf{x}^t \hat{\beta}(p_{h-1})\right]$$
$$\geq 0 \quad h = 2, \ldots, k \quad \forall \mathbf{x} \in S^m \tag{15}$$

where, S^m is the convex hull of a set of the n observed $\mathbf{x} \in R^m$. More specifically, S^m is the intersection of all convex sets containing the observations:

$$S^m \equiv \left\{ \sum_{i=1}^{b} \lambda_i \mathbf{x}_i \,\middle|\, \lambda_i \geq 0 \forall i, \; \sum_{i=1}^{b} \lambda_i = 1 \right\} \tag{16}$$

An obvious method to obtain the matrix $\mathbf{B}$ is the execution of an estimation procedure for each of the k different conditional quantile regressions. In the absence of further restrictions, the estimators to be included in $\mathbf{B}$ would be obtained by solving the minimization problem (4) 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.

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 covariate (with the exclusion of the first columns consisting entirely of ones). To simplify the evaluation of constraints (15), we can transform the covariates so that they range into the iinterval [0,1]:

$$v_{i,j} = \begin{cases} \dfrac{x_{i,j} - L_j}{U_j - L_j} & for \; j = 2, \cdots, m; \\ 1 & for \; j = 1 \end{cases} ; i = 1, 2, \cdots, n. \tag{17}$$

The domain of interest is now reduced from S^m to $D = \{\mathbf{v} | \mathbf{v} \in [0,1]^m\}$. The linear transformation (17) can easily be inverted after the estimation, while retaining the properties of the quantile regression estimators. In fact, if $\tilde{\beta}(p)$ is the vector of estimated parameters associated with $p \in P$ under (17), then:

$$\hat{\beta}(p) = \begin{cases} \dfrac{\tilde{\beta}_j(p)}{U_j - L_j} & for \; j = 2, \cdots, m; \\ \tilde{\beta}(p) - \sum_{j=2}^{m} \left[\dfrac{\tilde{\beta}_j(p) L_j}{U_j - L_j} \right] & Otherwise \end{cases} \tag{18}$$

Hence a quantile regression estimate of the unknown parameters is given by:

$$\mathbf{Q}(\tilde{\beta}, \mathbf{y}, \mathbf{v}) = \min_{\tilde{\beta} \in R^m} \left\{ \begin{array}{l} p \sum_{i | y_i \geq v_i^t \tilde{\beta}} (y_i - v_i^t \tilde{\beta}) + (1 - p) \\ \sum_{i | y_i \geq v_i^t \tilde{\beta}} (v_i^t \tilde{\beta} - y_i) \end{array} \right\}. \tag{19}$$

where $\mathbf{v}_i = (v_{i,1}, \cdots, v_{i,m})$. The linear programming theory assures that feasible solutions to (19) occur at the vertices of D, i.e., integers $M \subset (1, \cdots, n)$ corresponding to observations for which: $\tilde{\beta}(p) = [\mathbf{V}(M)]^{-1} \mathbf{y}(M)$, 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 (19) for $p \in P$ as follows:

$$\gamma(p_1) = \tilde{\beta}(p_1)$$
$$\gamma(p_j) = \tilde{\beta}(p_j) - \tilde{\beta}(p_{j-1}) \quad j = 2, \cdots, k. \tag{20}$$

The restrictions described in (15) are equivalent to:

$$\mathbf{v}_i^t \gamma(p_h) \geq 0 \quad \forall \mathbf{v} \in D \; h = 2, \cdots, k. \tag{21}$$

This condition, according to Bondell *et al.* (2010), is both necessary and sufficient to prevent overlapping hyperplanes. The merit of this approach is that the question of quantile crossings is now reduced to a linear programming problem, which can be solved via standard software.

Non-Crossing Weighted 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. Amerise (2016) for a good review.

Consistent with this premise, we believe that a way to avoid intersections between estimated hyperplanes (over the design space) is to put more emphasis on observations which are more coherent with the model (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 quantiles under the non-crossing restrictions (15) by adjusting fit of h-th quantile regression to the following objective function:

$$\mathbf{Q}\left(\tilde{\beta},\mathbf{y},\mathbf{v}\right)=\min_{\tilde{\beta}\in R^{m}}\left\{\begin{array}{c} p_{h}\sum_{i|y_{i}\ge\mathbf{v}_{i}^{t}\tilde{\beta}}w_{i,n}\left(y_{i}-\mathbf{v}_{i}^{t}\tilde{\beta}\right)+\left(1-p_{h}\right)\\ \sum_{i|y_{i}\ge\mathbf{v}_{i}^{t}\tilde{\beta}}w_{i,n}\left(\mathbf{v}_{i}^{t}\tilde{\beta}-y_{i}\right)\end{array}\right\}. \quad (22)$$

where, the weighting system **w** verifies the conditions.

$$w_{i,n}>0,\ \lim_{\mathbf{Q}\left(\tilde{\beta},\mathbf{y},\mathbf{v}\right)\to0}w_{i,n}=\omega>0\ \text{for}\ i=1,\ldots,n. \quad (23)$$

The magnitude of $w_{i;n}$ quantifies the suitability of the information contained in the i-th observation relatively to the k regression hyperplanes fitted to the n data points. Strictly positive weights are strongly recommended by Koenker (2013)[p. 17] since a null weight is ambiguous. Moreover, the system of weights tends to the equal weighting scheme when hyperplane do no cross. Note that the weighted version of the objective function (22) can be solved by applying the unweighted algorithm to the responses and design vectors defined by $r_{i,h}=w_{i,n}\,y_{i},\mathbf{z}_{i}=w_{i,n}\mathbf{v}_{i}\ i=1,2,\cdots,\ n$. Therefore, problem (22) can be reformulated as:

$$\mathbf{Q}\left(\tilde{\beta},\mathbf{r},\mathbf{z}\right)=\min_{\tilde{\beta}\in R^{m}}\left\{\begin{array}{c} p_{h}\sum_{i|r_{i}\ge\mathbf{z}_{i}^{t}\tilde{\beta}}\left(r_{i}-\mathbf{z}_{i}^{t}\tilde{\beta}\right)+\left(1-p_{h}\right)\\ \sum_{i|r_{i}\ge\mathbf{z}_{i}^{t}\tilde{\beta}}w_{i,n}\left(\mathbf{z}_{i}^{t}\tilde{\beta}-r_{i}\right)\end{array}\right\}. \quad (24)$$

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 source known to be polluted by errors. In practice, however, this assumption rarely holds so estimated weights must be used instead. There are many ways to estimate **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 $y_{i}-\hat{y}_{i}$ y_{i} where y_{i} is the i-th value of the response in a sample of n points and $\hat{y}_{i}$ is some estimate of $y_{i}|\mathbf{x}_{i}$ obtained with an unconstrained quantile regression hyperplane.

Let $\hat{\mathbf{C}}$ the ($n\times k$) matrix with columns given by the n estimated residuals of the noncrossing quantile regression associated with $p_{h,h}=1,\cdots,k$. This implies that the unweighted estimators of Bondell *et al.* (2010) are taken as a benchmark against which to compare weighted non-crossing quantile regressions. Furthermore, let $\hat{\mu}=\left(\hat{\mu}_{1},\cdots,\hat{\mu}_{k}\right)$ represent the ($k\times1$) vector of averages and $\hat{\Sigma}$ the ($k\times k$) matrix of variance-covariances of the k columns of $\hat{\mathbf{C}}$. It must be observed tha $\hat{\mu}_{j}=n^{-1}\sum_{i=1}^{n}\hat{c}_{i,j}$ is different from zero, unless, the distribution of conditional residuals is symmetrical. In order to quantify the potential impact on parameter estimation of $\hat{c}_{i}$ (the i-th row of $\hat{\mathbf{C}}$) we use the Mahalanobis distance between $\hat{c}_{i}$ and the null vector (the ideal perfect fit):

$$d_{i}=\sqrt{\hat{\mathbf{c}}_{i}\hat{\Sigma}^{-1}\hat{\mathbf{c}}_{i}}\ \ i=1,\cdots,n. \quad (25)$$

with $\hat{\sigma}_{h,h'}=\left(n-1\right)^{-1}\sum_{i=1}^{n}\left(\hat{c}_{i,h}-\hat{\mu}_{h}\right)\left(\hat{c}_{i,h'}-\hat{\mu}_{h'}\right)$. Naturally, there is a range of possibilities for converting distances into weights. An exponential transformation is especially appealing to us because of its simplicity and flexibility:

$$w_{i}=\exp\left\{-\tau d_{i}\right\}\ \ i=i,\cdots,n. \quad (26)$$

where $\tau\ge0$ is a finite-dimensional parameter that may be varied to modify the influence of the distances. Increasing values of t make the observation which is at distance one from the null vector progressively less relevant. For a given $\tau>0$, weights decrease as distances from the null vector increases. Form another point of view, if we consider the resemblance between (26) and the density function of an exponential random variables, then t can be thought to be the inverse of the expected uncertainty of the observations.

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 for onedimensional optimization offered in Base-R. The method used is a combination of golden section search and successive parabolic interpolation which searches a specified interval from lower to upper for a minimum of a function. A solution can be considered feasible if and only if $\mathbf{Q}\left(\beta^{*},\mathbf{r},\mathbf{z}\right)\le\mathbf{Q}\left(\tilde{\beta},\mathbf{r},\mathbf{z}\right)$ where β^{*} is the non-crossing weighted quantile estimate of the unknown parameters.

For what concerns large sample properties of non-crossing weighted quantile regressions, consider a set of percentages $p_{1}<p_{2}<\cdots,<p_{k}$ such that $p_{h}\in[\varepsilon,1-\varepsilon]$ for h = 1,···, k and $0<\varepsilon<0.5$ and assume:

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

$$a\le f_{yi|\mathbf{x}}\left[Q_{yi|\mathbf{x}}\left(p\right)\right]\le b;\qquad\left|f_{yi|\mathbf{x}}'\left[Q_{yi|\mathbf{x}}\left(p\right)\right]\right|\le c$$

uniformly for $x \in D$, $\varepsilon \le p \le (1-\varepsilon)$ and uniformly in $i = 1,\cdots, n$.

Under the above conditions, Bondell *et al*. (2010) prove that the estimator obtained via (22) is asymptotically equivalent to the typical 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.

Experimental Results

The examples and 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 algorithm for the WNC.
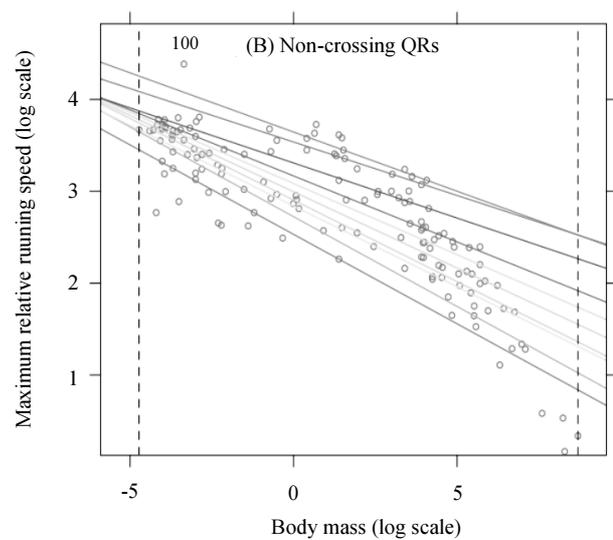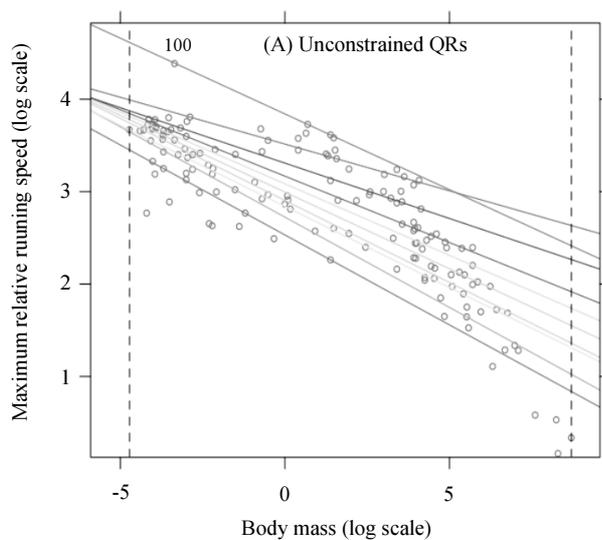
Iriarte-D´ıaz (2002) discusses the relationship between maximum relative running speed (body length/second) and body mass (kg) concerning $n = 142$ species of terrestrial mammals, in order to evaluate whether the relative locomotor performance shows a differential scaling depending on the range of mass analyzed. Overall, maximum relative running speed was found to decrease 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 of non-crossing quantiles proposed by Bondell *et al*. (2010) avoids the intersections which

are present in graph A, at least within the design domain delimited by the vertical lines. Our version with weights inversely related to the residuals of the quantile regressions (graph C) generates regression lines which not only bypass crossings, but also gather near the center of the observed points. It must be noted, in fact, that there is an entity which does not match the general impression: 100 corresponding to the heteromyid rodent (Dipodomys merriami). The bias attributable to the presence of this outlyier can be noticed looking at the highest two lines in graph A) and B. In the former there are crossings clearly due to the carry-over effect from the isolated point. In the latter, the problem of the crossing is solved, but the outlier shares the line with other regular entities. In graph C the influence of the outlier has been entirely 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 absolute errors affecting the various regression models. 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 our weighted version of the non-crossing regression quantiles attains a better performance than the standard procedure with respect the abso-lute residuals. The unweighted non-crossing technique does not improve, from a fitting point of view, upon unconstrained quantile regressions.

As the secnd 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 Fig. 2 it is shown that, in absence of outliers in the data and non-crossing lines, the three estimation methods behave similarly.
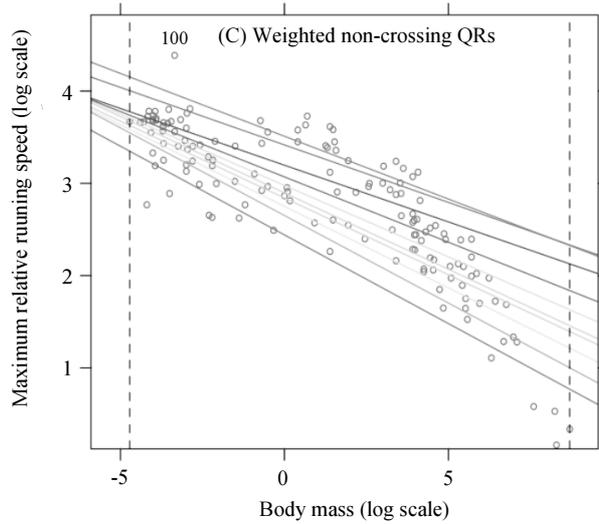

(A) Unconstrained QRs

(B) Non-crossing QRs

Fig. 1: Relationship between body mass and maximal relative running speed
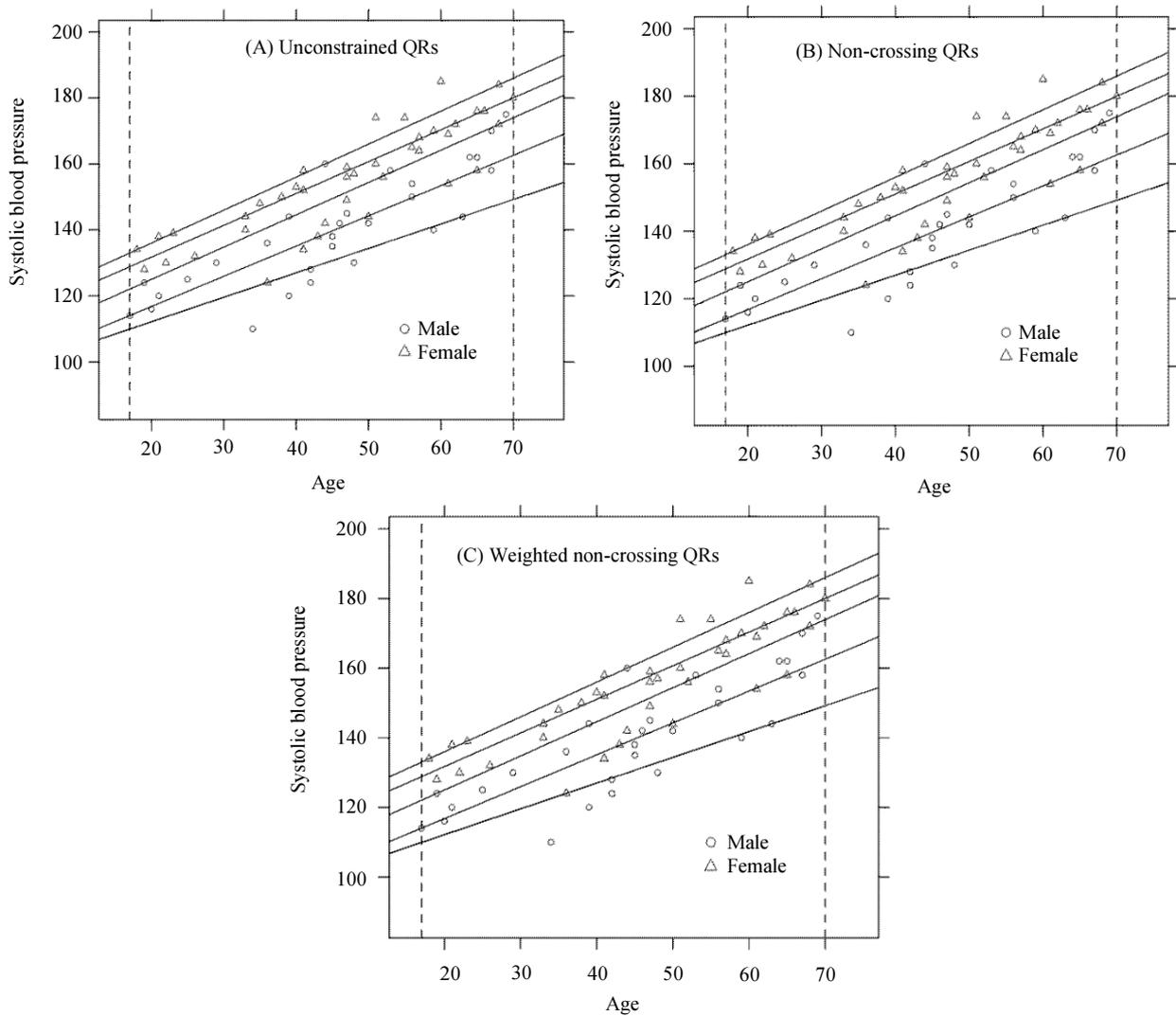

Fig. 2: Relationship between age and systolic blood pressure

115

Table 1: Fitting results of various estimation methods

	Unconstrained QR	Non-crossing QR	Weighted Nc QR
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

Table 2: Fitting results of various estimation methods

	Unconstrained QR			Non-crossing QR			Weighted Nc QR		
n	mean	max	min	mean	max	min	mean	max	min
120	4.7315	8.7818	2.9386	4.7131	8.3829	3.0049	3.6227	4.0412	3.3151
240	4.9565	9.3220	3.1068	4.9492	9.0667	3.1371	3.6846	4.3434	3.3514
360	4.9352	9.2070	3.1061	4.9285	9.1144	3.1196	3.8306	4.7738	3.3754
480	4.9422	9.2880	3.1007	4.9831	9.7505	3.1093	3.7369	4.4471	3.3318

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 14 non-constant predictor variables and one response variable, Corrected Median Value of owner-occupied homes (CMEDV). 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. 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.

Discussion and Conclusion

Conditional quantile functions ofer 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 (Parzen, 1979; 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 in sequence 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 $\hat{\beta}(p_h)$ has optimal properties for the *h*-th conditional distribution and this is also true for $h = 1, \cdots, k$. Since, the criterion (4) does not use the fact that $\beta(p_h), h = 1, \cdots, k$ lie in A^m, one might wonder if using such conditions in estimation procedures give a gain in efficiency. This is not necessarily so (Rothenberg, 1973[p. 55-57]) for the case of the linear least squares estimator). The question then is to find the best way of satisfying the constraints without worsening the properties 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.

Acknowledgement

The author would like to thank the reviewers for their constructive comments and suggestions.

Ethics

The Author declares there is not conflict of interest.

References

Amerise, I.L., 2016. Iteratively reweighted constrained quantile regressions. Adv. Applic. Statist., 49: 417-441. DOI: 10.17654/AS049060417

Bassett, G. and and R. Koenker, 1982. An empirical quantile function for linear models with iid errors. J. Am. Statist. Assoc. 77: 407-415. DOI: 10.1080/01621459.1982.10477826

Bassett, G. and R. Koenker, 1986. Strong consistency of regression quantiles and related empirical processes. Econometric Theory, 2; 191-201. DOI: 10.1017/S0266466600011488

Bondell, H.D., B.J. Reich and X. Wang, 2010. Noncrossing quantile regression curve estimation. Biometrika, 97: 825-838.

Buchinsky, M., 1998. Recent advances in quantile eegression models: A practical guideline for empirical research. J. Human Resources, 33: 88-126. DOI: 10.2307/146316

Dette, H. and S. Volgushev, 2008. Non-crossing non-parametric estimates of quantile curves. J. Royal Statist. Society, 70: 609-627. DOI: 10.1111/j.1467-9868.2008.00651.x

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

He, X., 1997. Quantile curves without crossing. Am. Statist., 51: 186-192. DOI: 10.1080/00031305.1997.10473959

Heagerty, P.J. and M.S. Pepe, 1999. Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. Applied Statist., 48: 533-551. DOI: 10.1111/1467-9876.00170

Iriarte-D´ıaz, J. 2002. Differential scaling of locomotor performance in small and large terrestrial mammals. J. Experimental Biol. 205: 2897-2908.

Koenker, R. 2013. Quantreg: Quantile regression. R package version 5.05.

Koenker, R. and G. Bassett, 1978. Regression quantiles. Econometrica, 46: 33-50. DOI: 10.2307/1913643

Koenker, R. and O. Geling, 2001. Reappraising medfly longevity: A quantile regression survival analysis. J. Am. Statist. Assoc., 96: 458-468. DOI: 10.1198/016214501753168172

Koenker, R. and V. D'Orey, 1987. Algorithm AS 229: computing regression quantiles. Applied Statist. 36: 383-393. DOI: 10.2307/2347802

Koenker, R. and V. D'Orey, 1994. Remark AS R92: A remark on algorithm AS 229: Computing dual regression quantiles and regression rank scores. Applied Statist., 43: 410-414. DOI: 10.2307/2986030

Koenker, R., 2015. Quantile Regression. 1st Edn., Cambridge University Press, New York, ISBN-10: 0521608279, pp: 349.

Liu, Y. and Y. Wu, 2011. Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. J. Nonparametr Statist., 23: 415-437. DOI: 10485252.2010.537336

Machado, J. and J. Mata, 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. J. Applied Econometr., 20: 445-465. DOI: 10.1002/jae.788

Melly, B., 2005. Decomposition of differences in distribution using quantile regression. Labour Econ., 12: 577-590. DOI: 10.1016/j.labeco.2005.05.006

Neocleous, T. and S. Portnoy, 2007. On monotonicity of regression quantile functions. Statist. Probab. Lett., 78: 1226-1229. DOI: 10.1016/j.spl.2007.11.024

Parzen, E. 1979. Nonparametric statistical data modelling. J. Am. Statist. Assoc., 74: 105-131. DOI: 10.1080/01621459.1979.10481621

Portnoy, S., 1991. Asymptotic behavior of the number of regression quantile breakpoints. SIAM J. Sci. Statist. Comput., 12: 867-883. DOI: 10.1137/0912047

Rothenberg, T.J., 1973. Efficient estimation with a priori information. 1st Edn., Yale University Press, New Haven, ISBN-10: 0300016077, pp: 180.

Schnabel, S. and P.H.C. Eilers, 2013a. A location-scale model for non-crossing expectile curves. Stat, 2: 171-183. DOI: 10.1002/sta4.27

Schnabel, S. and P.H.C. Eilers, 2013b. Simultaneous estimation of quantile curves using quantile sheets. Adv. Statist. Anal.

Seneta, E. and W.L. Steiger, 1984. A new lad curve-fitting algorithm: Slightly overdetermined equation systems in L_1. Discrete Applied Mathe. 7: 79-91. DOI: 10.1016/0166-218X(84)90115-X

Shim, J., C. Hwang and K.H. Seok, 2009. Non-crossing quantile regression via doubly penalized kernel machine. Comput. Statist., 24: 83-94. DOI: 10.1007/s00180-008-0123-y

Takeuchi, I. and T. Furuhashi, 2004. Non-crossing quantile regressions by SVM. Proceedings of the International Joint Conference on Neural Networks, Jul. 25-29, IEEE Xplore Press, Budapest, Hungary, pp: 401-406. DOI: 10.1109/IJCNN.2004.1379939

Takeuchi, I., Q.V. Le, T. Sears and A.J. Smola, 2006. Nonparametric quantile regression. J. Machine Learn. Res., 7: 1231-1264.

Tokdar, S.T. and J.B. Kadane, 2012. Simultaneous linear quantile regression: A semiparametric bayesian approach. Bayesian Anal., 7: 51-72. DOI: 10.1214/12-BA702

Wu, Y. and Y. Liu, 2009. Stepwise multiple quantile regression estimation using non-crossing constraints. Statist. Interface, 2: 299-310. DOI: 10.4310/SII.2009.v2.n3.a4

Yu, K. and M.C. Jones, 1998. Local linear quantile regression. J. Am. Statist. Assoc., 93: 228-237. DOI: 10.1080/01621459.1998.10474104

Zhao, Q., 2000. Restricted regression quantiles. J. Multivariate Anal., 72: 78-99. DOI: 10.1006/jmva.1999.1849