

Non-crossing quantile regression curve estimation

BY HOWARD D. BONDELL, BRIAN J. REICH, AND HUIXIA WANG

*Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695,
U.S.A.*

bondell@stat.ncsu.edu, reich@stat.ncsu.edu, wang@stat.ncsu.edu

SUMMARY

Since quantile regression curves are estimated individually, the quantile curves can cross, leading to an invalid distribution for the response. A simple constrained version of quantile regression is proposed to avoid the crossing problem for both linear and nonparametric quantile curves. A simulation study and a reanalysis of tropical cyclone intensity data shows the usefulness of the procedure. Asymptotic properties of the estimator are equivalent to the typical approach under standard conditions, and the proposed estimator reduces to the classical one if there is no crossing. The performance of the constrained estimator has shown significant improvement by adding smoothing and stability across the quantile levels.

Some key words: Crossing quantile curves; Heteroskedastic; Quantile regression; Robustness; Smoothing splines; Tropical cyclone.

1. INTRODUCTION

Quantile regression has become a useful tool to complement a typical least squares regression analysis (Koenker, 2005). Modeling of the median as opposed to the mean is much more robust to outlying observations. Additionally, examining the effect of the predictors on other quantiles can yield a clearer picture regarding the overall distribution of a response. In particular, in numerous instances, interest focuses on the effect of the predictors on the tails of a distribution, in addition to, or instead of, the center. Quantile regression is used to model the conditional quantiles of a response. Applications of quantile regression come from diverse areas, including economics, public health, meteorology, and surveillance.

However, when an investigator wishes to use quantile regression at multiple percentiles, the quantile curves can cross, leading to an invalid distribution for the response. In particular, given a set of covariates, it may turn out, for example, that the predicted 95th percentile of the response is smaller than the 90th percentile, which is impossible.

Consider a recent application of quantile regression to model tropical cyclone intensity (Jagger and Elsner, 2009). The goal is to model the maximum wind speed for near-coastal tropical cyclones occurring near the United States coastline based on climatological variables. Of particular interest is the upper tail of the distribution, as these are the storms that may cause excessive damage. Jagger and Elsner (2009) used quantile regression to examine the upper tail behavior as a function of four large-scale climate conditions. A sample of 422 tropical cyclones is used to model the maximum wind speed in terms of the climate covariates: the North Atlantic Oscillation Index, the Southern Oscillation Index, the Atlantic sea-surface temperature, and the average sunspot number.

49 With the focus being on the upper quantiles of the wind speed distribution, those corresponding
50 to category 4 and 5 hurricanes, consider using these data to fit a quantile regression at the set of
51 percentiles (0.25, 0.5, 0.75, 0.9, 0.95, 0.99). The fitted slopes of the quantile functions give the
52 effects of the covariates at the various levels of cyclone intensity. One particular issue with fitting
53 the upper quantiles is the lack of data, hence fitting individual quantile curves can be even more
54 problematic.

55 For example, if one were to examine the fitted quantiles for these data, the upper quantiles
56 cross not far from the mean. Due to the crossing, for instance, the practitioner would be forced to
57 claim that when North Atlantic Oscillation Index is one standard deviation below its mean and
58 the other three are one standard deviation above, the 90th percentile of the distribution of wind
59 speeds is larger than the 95th percentile. In addition, as further discussed in the analysis later,
60 inference regarding the significant predictors changes dramatically from one quantile to the next.
61 For example, predictors that appear highly significant in both the 90th percentile and the 99th
62 percentile may not appear significant at the 95th percentile. Although this is possible, much of it
63 can be attributed to the fact that the quantile functions are estimated separately.

64 This problem is well-known, but no simple and general solution currently exists. For the lin-
65 ear quantile regression model, Koenker (1984) considered parallel quantile planes to avoid the
66 crossing problem. Cole (1988) and Cole and Green (1992) assumed that a suitable transforma-
67 tion would yield normality of the response and proceeded to obtain nonparametric estimates of
68 the transformation along with the location and scale, which then fully determine the quantile
69 functions.

70 Similarly, He (1997) proposed a method to estimate the quantile curves while ensuring non-
71 crossing. However, the approach assumes a heteroskedastic regression model for the response,
72 which allows the predictors to affect the distribution of the response via a location and scale
73 change of an underlying base distribution. Although this is a flexible model, the predictors may
74 affect the response distribution in a less structured manner, which may not be captured by this
75 model. Furthermore, since the procedure is a sequential algorithm, the distributional properties of
76 the estimator are unclear. Simulation has also shown that even when the assumed heteroskedastic
77 model is correct, the estimation procedure does not necessarily improve upon the unconstrained
78 quantile regression estimator in finite samples. Wu and Liu (2009) have recently proposed an
79 algorithm to ensure non-crossing, by fitting the quantiles sequentially and constraining the cur-
80 rent curve to not cross the previous curve. One drawback of this algorithm is its dependence on
81 the order that the quantiles are fitted. Neocleous and Portnoy (2008) discuss interpolation of the
82 typical regression quantiles to ensure that, asymptotically, the probability of crossing will tend
83 to zero for the full quantile process.

84 The crossing problem persists for nonlinear quantile curves. Several authors have proposed to
85 first estimate the conditional cumulative distribution function via local weighting, and then invert
86 it to obtain the quantile curve. Hall, Wolff, and Yao (1999), Dette and Volgushev (2008), and
87 Chernozhukov, Fernandez-Val, and Galichon (2009) enforce the non-crossing via this approach
88 by modifying the estimate of the conditional distribution function. This indirect approach is
89 used if interest is purely in estimation of the conditional quantile. However, when interest also
90 focuses on quantifying the effects of the predictors, the quantile curves are typically modeled via
91 a parametric form, such as linear predictor effects, and a direct estimation approach is required.

92 In this paper, a direct correction to the quantile regression optimization problem is used to
93 ensure non-crossing quantile curves for any given sample. The approach is also extended to
94 nonparametric quantile curves.

2. NON-CROSSING QUANTILE REGRESSION

2.1. Alleviating the crossing problem

Let $x = (x_1, \dots, x_p)^T$, and denote $z = (1, x^T)^T$. Let $\mathcal{D} \subset R^p$, be a closed convex polytope, represented as the convex hull of N points in p -dimensions. Interest focuses on ensuring that the quantile curves do not cross for all values of the covariate $x \in \mathcal{D}$. Note that the covariate space is not assumed to be bounded, only the region of interest for non-crossing. If non-crossing were desired in linear quantile regression on an unbounded domain in any covariate direction, the result will be parallel lines, yielding the constant slope, location-shift model.

Assuming a linear quantile model, the τ th conditional quantile of the response is given by $z_i^T \beta_\tau$, i.e. $P(y_i \leq z_i^T \beta_\tau | x_i) = \tau$. The classical estimator of the regression coefficients for this quantile function is given by

$$\hat{\beta}_\tau = \operatorname{argmin}_\beta \sum_{i=1}^n \rho_\tau(y_i - z_i^T \beta), \quad (1)$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the so-called check function.

A typical quantile regression analysis will solve (1) separately for each of the q desired quantile levels, $\tau_1 < \dots < \tau_q$, to get $\hat{\beta}(\tau) = (\hat{\beta}_{\tau_1}^T, \dots, \hat{\beta}_{\tau_q}^T)^T$. Without any restriction, these resulting regression functions will often cross in finite samples, and hence the resulting conditional quantile curve for a given x will not be a monotonically increasing function of τ . Recall that $z = (1, x)^T$. Then, formally, non-monotonicity of the resulting estimated quantile function at a given point x is given by $z^T \hat{\beta}_{\tau_t} < z^T \hat{\beta}_{\tau_{t-1}}$ for at least one $t \in (2, \dots, q)$. For a simple example, it often turns out to be the case that the intercept is not an increasing function of τ , hence at $x = 0$ the quantile function is non-monotone. This becomes even more problematic with a larger number of predictors as the curves have a much larger space in which they may cross. When q increases, crossing also becomes much more likely.

To alleviate the crossing issue, it is proposed to estimate the quantiles simultaneously under the non-crossing restriction. Specifically, the optimization problem becomes

$$\begin{aligned} \hat{\beta}(\tau) = \operatorname{argmin}_\beta & \sum_{t=1}^q w(\tau_t) \sum_{i=1}^n \rho_{\tau_t}(y_i - z_i^T \beta_{\tau_t}) \\ \text{subject to } & z^T \beta_{\tau_t} \geq z^T \beta_{\tau_{t-1}} \text{ for all } x \in \mathcal{D} \text{ and all } t = 2, \dots, q, \end{aligned} \quad (2)$$

for some weight function, $w(\tau_t)$, such that $w(\tau_t) > 0$ for all $t = 1, \dots, q$.

Without the restriction, the solution to (2) is exactly the solution to (1) regardless of the choice of weight function $w(\tau_t)$. In fact, a direct consequence of this formulation yields that if the classical estimator obeys the non-crossing constraint for a given data set, the resulting estimator will be exactly the classical estimator. Furthermore, the asymptotic distribution of the estimator, as discussed in the next section, will not depend on the choice of weight function. A convenient practical choice of weight function is to equally weight the terms, i.e. $w(\tau_t) = 1$ for all t . This is the choice used in the examples.

Since the domain is the convex hull of N points, it suffices to enforce the non-crossing restriction at each vertex. Letting (z_1, \dots, z_N) denote the set of vertices, then any point in the region can be expressed as $\sum_{k=1}^N c_k z_k$ with $\sum_{k=1}^N c_k = 1$ and each $c_k \geq 0$. If $z_k^T \beta_{\tau_t} \geq z_k^T \beta_{\tau_{t-1}}$ for all $k = 1, \dots, N$, it follows that $\sum_{k=1}^N c_k z_k^T \beta_{\tau_t} \geq \sum_{k=1}^N c_k z_k^T \beta_{\tau_{t-1}}$. This can then be solved via standard linear programming with $N * (q - 1)$ linear constraints, where N is the number of vertices, and q is the number of quantiles. This may result in a large number of constraints.

However, it will now be shown that if the domain of interest can be reduced to $[0, 1]^p$, constraints at each of the 2^p vertices are unnecessary, and only a total of $q - 1$ constraints are needed,

rather than $N * (q - 1)$. Any domain of interest for which there exists an invertible affine transformation that maps to $[0, 1]^p$ can be used, so that the transformation is performed, and then transformed back after the estimation, while retaining the non-crossing property. This has the potential to simplify computation a great deal. Hence, one may wish to approximate the desired domain by one of this form. The remainder of the article will focus on the domain $\mathcal{D} = [0, 1]^p$.

To ensure non-crossing everywhere, consider the transformation from $\beta_{\tau_1}, \dots, \beta_{\tau_q}$ to $\gamma_{\tau_1}, \dots, \gamma_{\tau_q}$, where $\gamma_{\tau_1} = (\gamma_{0,\tau_1}, \gamma_{1,\tau_1}, \dots, \gamma_{p,\tau_1})^T = \beta_{\tau_1}$ and $\gamma_{\tau_t} = (\gamma_{0,\tau_t}, \gamma_{1,\tau_t}, \dots, \gamma_{p,\tau_t})^T = \beta_{\tau_t} - \beta_{\tau_{t-1}}$ for $t = 2, \dots, q$.

The constraint in (2) is equivalent to $z^T \gamma_{\tau_t} \geq 0$ for all $x \in \mathcal{D}$ and all $t = 2, \dots, q$. Break each γ_{j,τ_t} into its positive and negative parts, so that $\gamma_{j,\tau_t} = \gamma_{j,\tau_t}^+ - \gamma_{j,\tau_t}^-$, where both γ_{j,τ_t}^+ and γ_{j,τ_t}^- are non-negative, and only one may be non-zero. With this reparameterization, non-crossing can be easily ensured on all of $\mathcal{D} = [0, 1]^p$. Using this parameterization, the constraint in (2) becomes simply that

$$\gamma_{0,\tau_t} - \sum_{j=1}^p \gamma_{j,\tau_t}^- \geq 0 \quad (3)$$

for each $t = 2, \dots, q$.

This gives the point that is the worst case scenario for each t , having $x_j = 1$ when $\gamma_{j,\tau_t} < 0$, and $x_j = 0$ when $\gamma_{j,\tau_t} > 0$. Since this point is in \mathcal{D} , non-crossing must be enforced there, and hence (3) is a necessary condition. Since this point is the worst case, all other points in $\mathcal{D} = [0, 1]^p$ will then automatically satisfy the constraint for each $t = 2, \dots, q$, and hence (3) is also a sufficient condition.

After reparameterization, the problem is thus reduced to a straightforward linear programming problem, which can be solved via standard software. The linear program is extremely sparse and thus the use of a sparse matrix representation is more efficient. For computation, the linear programming has been implemented via use of the SparseM package (Koenker and Ng, 2003) in the R software platform (R Development Core Team, 2009), and code is available from the corresponding author.

2.2. Asymptotic properties

Consider a set of percentiles $\tau_1 < \dots < \tau_q$, such that $\tau_t \in [\varepsilon, 1 - \varepsilon]$ for all t , with $0 < \varepsilon < 1/2$. For the asymptotic properties the set (τ_1, \dots, τ_q) is allowed to change with n , if desired. In particular one may wish to consider a denser set as the sample size increases. Assume that the linear quantile regression model holds with true parameter $\beta^0(\tau)$, so that the τ_t^{th} quantile of the conditional distribution for the response is given by $z^T \beta_{\tau_t}^0$. Specifically, $F_{Y_i|x}^{-1}(\tau_t) = z^T \beta_{\tau_t}^0$ for $i = 1, \dots, n$, where $F_{Y_i|x}$ denotes the conditional distribution function for observation i . Let $\tilde{\beta}(\tau)$ be the classical quantile regression estimator obtained by solving (1) separately for each τ_t , and let $\hat{\beta}(\tau)$ be the constrained non-crossing version obtained via (2).

The following theorem shows that, regardless of choice of weight function, the estimator obtained via (2) is asymptotically equivalent to the typical quantile regression estimator. The following conditions are assumed.

1. The weights $w(\tau_t) > 0$ for all $t = 1, \dots, q$.
2. The matrix $n^{-1} \sum_{i=1}^n x_i x_i^T$ is positive definite.
3. The conditional distributions have densities, $f_{Y_i|x}$, that are differentiable with respect to Y_i for every x and each $i = 1, \dots, n$.
4. There exist constants $a > 0$, $b < \infty$, and $c < \infty$ such that

$$a \leq f_{Y_i|x} \left\{ F_{Y_i|x}^{-1}(\tau) \right\} \leq b \text{ and } \left| f_{Y_i|x}^{(1)} \left\{ F_{Y_i|x}^{-1}(\tau) \right\} \right| \leq c,$$

uniformly for $x \in \mathcal{D}$, $\varepsilon \leq \tau \leq 1 - \varepsilon$, and $i = 1, \dots, n$, where $f_{Y_i|x}^{(1)}$ denotes the derivative of $f_{Y_i|x}$ with respect to Y_i .

The first condition ensures that the chosen weight function is appropriate to estimate the desired quantile curves. The remaining conditions allow for a uniform Bahadur representation of the classical quantile regression estimator, as in Neocleous and Portnoy (2008), which ensures that $\tilde{\beta}_\tau - \beta_\tau^0 = O_p(n^{-1/2})$ uniformly in $\varepsilon \leq \tau \leq 1 - \varepsilon$.

THEOREM 1. *Let $\hat{\beta}(\tau)$ and $\tilde{\beta}(\tau)$ be the constrained and unconstrained estimators, respectively, for the set of quantiles $\tau_1 < \dots < \tau_q$, such that $n^{1/2} \min_t(\tau_{t+1} - \tau_t) \rightarrow \infty$. Then for any $u \in \mathfrak{R}^{pq}$,*

$$\left| P \left[n^{1/2} \left\{ \hat{\beta}(\tau) - \beta^0(\tau) \right\} \leq u \right] - P \left[n^{1/2} \left\{ \tilde{\beta}(\tau) - \beta^0(\tau) \right\} \leq u \right] \right| \rightarrow 0,$$

so that the constrained estimator has the same limiting distribution as the classical quantile regression estimator.

Based on Theorem 1, inference for the $n^{1/2}$ -consistent constrained quantile regression can be achieved by using the known asymptotic results for classical quantile regression. In particular, appropriate asymptotic standard errors can be computed via the quantile regression sandwich formula (Koenker, 2005, sec. 3.2.3).

3. EXTENSION TO NONPARAMETRIC QUANTILE CURVES

3.1. Quantile curves

Consider the model with a single predictor $x \in [0, 1]$. Without assuming that the quantiles vary linearly with x , a nonparametric fit is often done via quantile smoothing splines. When the quantiles are now curves, the crossing problem becomes even more pronounced, as the curves are more likely to cross.

Taking the approach of quantile smoothing splines (Koenker, Ng, and Portnoy, 1994), the constrained joint quantile smoothing spline estimate can be formulated as the set of functions $\hat{g}_{\tau_1}, \dots, \hat{g}_{\tau_q} \in \mathcal{G}$ that minimizes

$$\begin{aligned} & \sum_{t=1}^q w(\tau_t) \sum_{i=1}^n \rho_{\tau_t} \{y_i - g_{\tau_t}(x_i)\} + \sum_{t=1}^q \lambda_{\tau_t} V(g'_{\tau_t}) \\ & \text{subject to } g_{\tau_t}(x) \geq g_{\tau_{t-1}}(x) \text{ for all } x \in [0, 1] \text{ and all } t = 2, \dots, q, \end{aligned} \quad (4)$$

where $V(g')$ is the total variation of the derivative of the function g . For twice continuously differentiable g , the total variation of the derivative is given by $V(g') = \int_0^1 |g''(x)| dx$. In general,

$$V(g') = \sup_P \sum_{i=0}^{N_P-1} |g'(x_{i+1}) - g'(x_i)| dx, \quad (5)$$

where the supremum is taken over the set of all possible partitions P of $[0, 1]$, and N_P denotes the number of endpoints that defines the partition P .

Following Pinkus (1988) and Koenker et al. (1994), consider the expanded 2nd-order Sobolev space,

$$\mathcal{G} = \left\{ g : g(x) = a_0 + a_1 x + \int_0^1 (x - y)_+ d\mu(y), V(\mu) < \infty, a_i \in \mathfrak{R}, i = 0, 1 \right\},$$

241 where μ is a measure with finite total variation. This space includes the usual Sobolev space of
 242 functions having 2nd derivative with finite L_1 norm and absolutely continuous 1st derivative,
 243 while also including the limiting piecewise linear functions. This expansion is necessary, as
 244 discussed in Pinkus (1988), to ensure that the interpolating function that minimizes the total
 245 variation resides in the function space \mathcal{G} . For piecewise linear functions, the supremum in (5)
 246 occurs when the partition coincides with the breakpoints of the function.

247
 248 **THEOREM 2.** *The set of functions $\hat{g}_{\tau_1}, \dots, \hat{g}_{\tau_q} \in \mathcal{G}$ minimizing (4) subject to the non-crossing*
 249 *constraint, consists of non-crossing linear splines with knots at the data points.*

250
 251 This implies that it suffices to consider the problem in terms of a linear spline basis, for which
 252 the total variation is a linear function of the coefficients. Hence, as in Koenker et al. (1994),
 253 the smoothing spline problem is a linear programming problem. It will now be shown that the
 254 non-crossing quantile constraint can also be directly incorporated into this framework while still
 255 retaining the linear programming problem.

256 Let $\{B_j(x)\}$ for $j = 1, \dots, k_n + 1$ denote the linear B-spline basis with k_n internal knots and
 257 endpoints at 0 and 1. Then $\hat{g}_{\tau_t}(x) = \hat{\beta}_{0,\tau_t} + \sum_{j=1}^{k_n+1} \hat{\beta}_{j,\tau_t} B_j(x)$. Using the analogous parame-
 258 terization as in the previous section in terms of differences in the coefficients across quantile
 259 levels, i.e. $\gamma_{\tau_1} = \beta_{\tau_1}$ and $\gamma_{\tau_t} = \beta_{\tau_t} - \beta_{\tau_{t-1}}$ for $t = 2, \dots, q$, it follows that $\hat{g}_{\tau_t}(x) - \hat{g}_{\tau_{t-1}}(x) =$
 260 $\hat{\gamma}_{0,\tau_t} + \sum_{j=1}^{k_n+1} \hat{\gamma}_{j,\tau_t} B_j(x)$ for any x . Hence the differences across quantiles is simply written in
 261 terms of a linear B-spline basis.

262 Considering the differences across successive quantiles as a linear spline with knots at each
 263 data point, it follows that it is necessary and sufficient to enforce the non-negative constraint at
 264 the knots. Non-negativity at each knot, will imply non-negativity between the knots, due to the
 265 linearity. The form of the linear B-spline basis allows for a convenient parametrization, since at
 266 each knot, a single basis function takes the value 1, while the remaining are 0. Hence at x_j , where
 267 x_j is the value at knot j , the difference across the successive quantiles is given by $\hat{g}_{\tau_t}(x_j) -$
 268 $\hat{g}_{\tau_{t-1}}(x_j) = \hat{\gamma}_{0,\tau_t} + \hat{\gamma}_{j,\tau_t}$. Hence, letting $\gamma_{j,\tau_t} = \gamma_{j,\tau_t}^+ - \gamma_{j,\tau_t}^-$ be the coefficients parameterized as
 269 above, it is necessary and sufficient to constrain $\gamma_{0,\tau_t} - \max_j(\gamma_{j,\tau_t}^-) \geq 0$ for each $t = 2, \dots, q$.
 270 This can now be turned into a linear programming problem via using the set of constraints $\gamma_{0,\tau_t} -$
 271 $\gamma_{j,\tau_t}^- \geq 0$ for each $t = 2, \dots, q$ and $j = 1, \dots, k_n + 1$.

272 Using the linear B-spline basis, the total variation penalty is simply a linear function of the
 273 basis coefficients. Hence, the full optimization problem given by (4) is again a linear program-
 274 ming problem, and computation can be done efficiently using the sparse matrix representation as
 275 before.

277 3.2. Tuning the procedure

278 Koenker et al. (1994) and He and Ng (1999) suggest the use of a Schwartz-type information
 279 criterion (SIC) for choosing the regularization parameter in quantile smoothing splines. For each
 280 individual quantile curve, the SIC criterion is given by

$$282 \text{SIC}(\lambda_{\tau_t}) = \log \left[n^{-1} \sum_{i=1}^n \rho_{\tau_t} \left\{ y_i - \hat{g}_{\tau_t}^{\lambda_{\tau_t}}(x_i) \right\} \right] + (2n)^{-1} p_{\lambda_{\tau_t}} \log(n), \quad (6)$$

283
 284
 285
 286 where $\hat{g}_{\tau_t}^{\lambda_{\tau_t}}$ denotes the estimated function for that choice of λ_{τ_t} , while $p_{\lambda_{\tau_t}}$ is the number of
 287 interpolated data points, i.e. those with zero residuals, and serves as a natural measure of the
 288 complexity of the model. The full set of tuning parameters for the individual quantile curves

minimizes the joint Schwartz-type criterion

$$\text{SIC}_J(\lambda) = \sum_{t=1}^q w_*(\tau_t) \log \left[n^{-1} \sum_{i=1}^n \rho_{\tau_t} \left\{ y_i - \hat{g}_{\tau_t}^{\lambda_{\tau_t}}(x_i) \right\} \right] + (2n)^{-1} \log(n) \sum_{t=1}^q p_{\lambda_{\tau_t}}, \quad (7)$$

for any choice of weights $w_*(\tau_t) > 0$. The weights, $w_*(\tau_t)$, may differ than those in the loss function due to the log scale. However, in the case of $w(\tau_t) = 1$, as in the examples, it is natural to also choose $w_*(\tau_t) = 1$.

For the proposed joint non-crossing quantile smoothing spline, the joint Schwartz-type criterion (7) is used to select the tuning parameters. If the individual quantile smoothing splines chosen via the separate SIC criteria (6) do not cross, then the joint quantile smoothing spline will be exactly this set.

One could instead choose the logarithm of the full objective function, i.e. taking the logarithm outside the double summation, instead of summing up the individual pieces. However, the proposed criterion in (7) will guarantee that the individual quantile smoothing splines agree with the joint non-crossing smoothing spline in the case that the former are non-crossing. The alternative criterion will not necessarily maintain this desirable property. In addition, simulation results have shown that the criterion given by (7) seems to exhibit better empirical performance.

In the joint quantile smoothing spline, there are separate tuning parameters for each quantile curve, resulting in a q -dimensional tuning parameter selection problem. However, experience, as shown in the simulations, has shown that a single tuning parameter performs sufficiently well in controlling the overall smoothness of the set of curves, while still allowing for varying degrees of smoothness among the curves themselves.

If a full q -dimensional tuning is desired, a directed search can proceed as follows. First fit the joint smoothing spline with a single tuning parameter, i.e. setting $\lambda_{\tau_1} = \dots = \lambda_{\tau_q}$ and find the value that minimizes the SIC_J criterion. Next vary λ_{τ_1} on a grid around the current value and minimize the SIC_J criterion keeping the remaining λ_{τ_t} values fixed at the current value. Take this as the new tuning parameter vector. Continue to update sequentially in this fashion, until no further improvement is possible. While this algorithm is not guaranteed to find the global minimizer, it will at least converge to a local minimizer, starting from the optimal solution for the single tuning parameter case.

4. SIMULATION STUDY

4.1. Linear quantile regression

The proposed method is now compared to classical quantile regression without the non-crossing assumption. Additionally, the method is compared to the approach of He (1997) which assumes a location-scale heteroskedastic error model. Each of the examples are a special case of the heteroskedastic error model as in He (1997)

$$y_i = \beta_0 + \beta^T x_i + \left(\gamma_0 + \gamma^T x_i \right) \varepsilon_i, \quad x_{ij} \sim U(0, 1), \quad \varepsilon_i \sim N(0, 1). \quad (8)$$

For all examples, the intercepts are set to $\beta_0 = \gamma_0 = 1$. In each example, 6 quantile curves, $\tau = 0.1, 0.3, 0.5, 0.7, 0.9, 0.99$, are fitted to the data, either separately for the classical quantile regression approach, or simultaneously for the proposed approach. The method of He (1997) estimates the location and scale parameters in the model directly, and then uses them to estimate the quantile curves. For each example 500 data sets are simulated. The three examples are given as

337 *Example 1:* The sample size is $n = 100$, with $p = 4$ predictors, and parameters $\beta =$
 338 $(1, 1, 1, 1)^T$, and $\gamma = (0.1, 0.1, 0.1, 0.1)^T$.

339 *Example 2:* The sample size is $n = 100$, with $p = 10$ predictors, and parameters $\beta =$
 340 $(1, 1, 1, 1, 0^T)^T$, and $\gamma = (0.1, 0.1, 0.1, 0.1, 0^T)^T$.

341 *Example 3:* The sample size is varied in $n = (100, 200, 500)$, with $p = 7$ predictors, and pa-
 342 rameters $\beta = (1, 1, 1, 1, 1, 1, 1)^T$ and $\gamma = (1, 1, 1, 0, 0, 0, 0)^T$.

343 To illustrate the crossing problem, for the first example, out of 500 generated data sets, 491 of
 344 them had at least one crossing somewhere in the domain. To examine the effect of increasing the
 345 sample size, the third example was run with $n = 100, 200$, and 500.

346 The results from the examples are given in Table 1. Presented are the empirical root mean
 347 integrated squared errors in estimation of the curves for each of $\tau = 0.5, 0.9, 0.99$ given by
 348 $RMISE = \left[n^{-1} \sum_{i=1}^n \{ \hat{g}_\tau(x_i) - g_\tau(x_i) \}^2 \right]^{1/2}$, where \hat{g}_τ is the estimated function and g_τ is
 349 the true function. In this case, the functions g and \hat{g} are linear, while in the next section they are
 350 non-linear. The table presents the average root mean integrated squared errors over the 500 data
 351 sets along with its estimated standard error. The results for the other quantiles are similar and
 352 thus omitted. However, by fitting all simultaneously, it is guaranteed that they will not cross.

353 In each of the settings considered, the proposed constrained approach gives significantly better
 354 estimation for all quantiles. This is probably because the constraints add some smoothness and
 355 stability across the quantile curves. Since the true curves do not cross, it is expected that the
 356 constrained estimates would perform better. This improvement is greater in the tails, as there is
 357 less data, hence the smoothing effect of the constraint allows for the borrowing of strength. In
 358 addition, the improvement is much more pronounced for example 2, due to the presence of the
 359 extra irrelevant predictors.

360 By varying the sample size in example 3, as expected, the differences become smaller as
 361 the sample size grows, due to the consistency of the classical estimators. However, even with
 362 $n = 500$ some statistically significant improvement still remains, as seen by the differences in
 363 root mean integrated squared errors relative to their reported standard errors. The estimator of He
 364 (1997), although based on the heteroskedastic model, does not exhibit better performance than
 365 the typical estimator except in the extreme quantiles in the larger sample case. This phenomenon
 366 was also observed by Wu and Liu (2009), and is probably because it requires implicit nonpara-
 367 metric estimation of the quantiles of the residuals, which may be unstable for smaller sample
 368 sizes.

369 Examples with other setups, including covariates generated from gaussian distributions, gave
 370 similar results, and are thus omitted.

371 4.2. Nonparametric quantile regression

372 For nonparametric quantile regression, the proposed method is compared with quantile
 373 smoothing splines and quantile regression splines. Quantile smoothing splines results in penal-
 374 ized linear splines, as in the proposed method, but with each curve fitted independently. Quantile
 375 regression splines starts with linear splines and performs knot selection. Both methods are im-
 376 plemented in the R package COBS (Constrained B-Spline Smoothing, He and Ng, 1999; Ng and
 377 Maechler, 2007). This package also allows the user to specify qualitative constraints on each
 378 individual quantile curve, such as monotonicity or convexity. This constraint on the curves refers
 379 to constraining the quantile curve for a particular value of τ , not as a function across values of τ
 380 as would be needed to ensure non-crossing. To ease the computational burden, the COBS pack-
 381 age, by default, implements the quantile smoothing splines with 25 knots instead of knots at each
 382 data point. The user may specify more or less knots, if desired. The same choice of a reduced
 383
 384

385 set of knots is used for computational convenience for the proposed joint smoothing spline in the
 386 simulation study.

387 Each of the two examples are again special cases of the heteroskedastic error model, given by

$$388 \quad y_i = f(x_i) + g(x_i) \varepsilon_i, \quad (9)$$

389
 390 for some functions f and g . The covariate is again generated as $U(0, 1)$ and $\varepsilon_i \sim N(0, 1)$ with
 391 $n = 100$. The two examples are given as

392
 393 *Example 4:* The mean function is $f(x) = 0.5 + 2x + \sin(2\pi x - 0.5)$, and the variance func-
 394 tion is $g(x) = 1$.

395 *Example 5:* The mean function is $f(x) = 3x$, and the variance function is $g(x) = 0.5 + 2x +$
 396 $\sin(2\pi x - 0.5)$.

397
 398
 399 Example 4 results in the quantile curves simply being a shift in the intercept as shown in
 400 the left panel of Figure 1. Example 5 results in the quantile curves having various degrees of
 401 smoothness, with the median being linear, and more curvature exhibited in the extreme quantile
 402 curves, as shown in the right panel of Figure 1.

403 For the nonparametric fits, the 0.99 quantile was not used, as it is too extreme for a
 404 nonparametric estimate based on a sample of size 100. Hence the set of quantiles, $\tau =$
 405 $0.1, 0.3, 0.5, 0.7, 0.9$ was fitted. Table 2 shows the results for $\tau = 0.5, 0.7, 0.9$ for the two ex-
 406 amples over the 500 simulated data sets for each example. The lower quantiles are analogous
 407 due to symmetry. Overall, the proposed method compares favorably to the traditional quantile
 408 splines, in terms of integrated squared error. For comparison, the proposed method is computed
 409 using a single tuning parameter as well as using a separate tuning parameter for each quantile
 410 level. In both examples, the single tuning parameter exhibits very similar performance to the full
 411 q -dimensional tuning.

412 413 4.3. Varying the choice of quantiles

414 Assume the linear quantile regression model holds for a given quantile of interest. Since the
 415 proposed non-crossing approach is based on simultaneous estimation of a set of quantile curves,
 416 the estimate for the quantile of interest will change depending on the included quantiles. For
 417 example, if interest focuses on the median, one can add any number of additional quantiles along
 418 with the median. Based on the results of Theorem 1, the asymptotic distribution will not be
 419 affected by the number of quantiles added. However, adding additional quantiles can improve
 420 the finite sample results via adding stability to the estimation.

421 Figure 2 plots mean squared error in estimating the slope at the median for a univariate model
 422 as a function of the number of included quantiles for each of sample sizes $n = 50, 100,$ and 200 .
 423 The median regression estimate was computed on 5000 datasets based on an increasingly dense
 424 sequence of equally spaced quantiles. The y-axis represents the ratio of mean squared error to
 425 that of using only the median. The use of more quantiles seems to help stabilize the estimation.

426 It is natural to assume that if all quantiles are linear, adding more would give better results. Fig-
 427 ure 2 was generated via the model in example 5, so that only the median is linear. However, even
 428 with the misspecification, there is still a gain in accuracy, which is probably because the quan-
 429 tiles near the median are still close to linear, and help stabilize the estimation. This phenomenon
 430 was observed in other scenarios, including those focused on extreme quantiles. In practice, we
 431 recommend adding quantiles in a neighborhood of the quantile of interest until the estimation
 432 appears to stabilize.

5. ANALYSIS OF HURRICANE DATA

The non-crossing quantile regression approach is now applied to the tropical cyclone data. The data consists of a sample of 422 tropical cyclones occurring near the US coastline over the period 1899-2006. Jagger and Elsner (2009) used linear quantile regression to model the maximum wind speed of each cyclone as a function of four climate covariates: the North Atlantic Oscillation Index, the Southern Oscillation Index, the Atlantic sea-surface temperature, and the average sunspot number. The climate covariates are constant within a single year to represent the yearly large-scale climate conditions. The North Atlantic Oscillation Index is the pre-season and early season average of the May and June values. The other three are obtained by averaging over the peak season of August through October. The particular focus is on the upper quantiles, as these extreme hurricane-strength storms are of considerable importance.

Following Jagger and Elsner (2009), quantile regression is applied to these data at the quantiles 0.25, 0.5, 0.75, 0.9, 0.95, 0.99. Table 3 shows the parameter estimates for the intercept along with the four slope parameters for the three upper quantiles from both a classical quantile regression fit and the proposed non-crossing fit. As in Jagger and Elsner (2009), included are pointwise 90% confidence intervals. In the other three quantiles, the results for both methods are similar.

Of particular note is the smoothing of the coefficients in the extreme quantiles. This smoothing stabilizes the inference and avoids some of the possibly spurious associations, such as the significance of the sunspot number in the 99th percentile but not in any of the other upper quantiles. The confidence intervals for both methods were obtained using the asymptotic normality and using the kernel method to estimate the inverse density needed for the standard errors (Powell, 1991; Koenker, 2005).

As the estimates will change depending on the number and location of included quantiles, a sensitivity analysis is performed. The coefficient estimates for the median and 0.99 quantile regressions are examined as a function of an increasing number of quantiles. Figure 3 plots the estimated coefficients for each of the four slopes for the median (left panel) and the 0.99 quantile (right panel). Initially, only the two quantiles were fitted. Then quantiles were sequentially added until a grid spacing of 0.05 was obtained, with a grid spacing of 0.01 bracketing the median (from 0.45 to 0.55) and a spacing of 0.01 from 0.9 to 0.99, to ensure a more saturated region around the quantiles of interest. The median is much less sensitive, while the 0.99 quantile is clearly highly sensitive. Once 14 quantiles are included, the inference regarding significant predictors remains the same the rest of the way, and is the same as that reported in Table 3. This is at the point in the sensitivity analysis that the 0.95 quantile is first added between the 0.9 and the 0.99 quantiles.

ACKNOWLEDGEMENT

The authors are grateful to the editor, an associate editor, and two anonymous referees for their valuable comments. This research was sponsored by the National Science Foundation, U.S.A., and the National Institutes of Health, U.S.A.

APPENDIX

Proof of Theorem 1

Let \hat{Z}_n and \tilde{Z}_n denote $n^{1/2} \{ \hat{\beta}(\tau) - \beta^0(\tau) \}$ and $n^{1/2} \{ \tilde{\beta}(\tau) - \beta^0(\tau) \}$, respectively. Then

$$\left| P \left(\hat{Z}_n \leq u \right) - P \left(\tilde{Z}_n \leq u \right) \right| = \left| P \left(\hat{Z}_n \leq u \mid \hat{Z}_n \neq \tilde{Z}_n \right) - P \left(\tilde{Z}_n \leq u \mid \hat{Z}_n \neq \tilde{Z}_n \right) \right| P \left(\hat{Z}_n \neq \tilde{Z}_n \right).$$

481 Since the first term in the product is bounded by 1, it suffices to show that $P\left(\hat{Z}_n \neq \tilde{Z}_n\right) \rightarrow 0$,
 482 or $P\left(\hat{Z}_n = \tilde{Z}_n\right) \rightarrow 1$.
 483

484 Due to the formulation of the estimator, the event $\hat{Z}_n = \tilde{Z}_n$ is equivalent to the event that the
 485 classical quantile regression estimator maintains its appropriate ordering. To show that the prob-
 486 ability of this event goes to one, consider the difference in the classical estimator at successive
 487 quantiles, $n^{1/2}(z^T \tilde{\beta}_{\tau_{t+1}} - z^T \tilde{\beta}_{\tau_t})$. It will be shown that this difference must be positive with
 488 probability tending to one for every $t = 1, \dots, q$.

489 The difference can be written as

$$490 \quad n^{1/2}(z^T \tilde{\beta}_{\tau_{t+1}} - z^T \beta_{\tau_{t+1}}^0) - n^{1/2}(z^T \tilde{\beta}_{\tau_t} - z^T \beta_{\tau_t}^0) + n^{1/2}(z^T \beta_{\tau_{t+1}}^0 - z^T \beta_{\tau_t}^0).$$

492 Under the assumed regularity conditions, the classical quantile regression estimator, $\tilde{\beta}(\tau)$, is
 493 $n^{1/2}$ -consistent. Hence, it follows that the first two terms are $O_p(1)$ for any t .

494 By the Mean Value Theorem, it follows that $z^T \beta_{\tau_{t+1}}^0 - z^T \beta_{\tau_t}^0 = (\tau_{t+1} - \tau_t) \frac{\partial}{\partial \tau} z^T \beta_{\tau}^0|_{\tau=\tau^*}$,
 495 where $\tau_t \leq \tau^* \leq \tau_{t+1}$. Now, regularity condition (3) yields that $\frac{\partial}{\partial \tau} z^T \beta_{\tau}^0 =$
 496 $\left\{f_{Y_i|x}\left(F_{Y_i|x}^{-1}(\tau)\right)\right\}^{-1} \geq 1/b$ for any $\tau \in (0, 1)$. Hence $n^{1/2}(z^T \beta_{\tau_{t+1}}^0 - z^T \beta_{\tau_t}^0) \geq$
 497 $n^{1/2}(\tau_{t+1} - \tau_t)/b$. By assumption, the right hand side diverges. This leads to the third
 498 term dominating in the difference with probability tending to one, and thus the difference will
 499 be positive.
 500

501
 502 *Proof of Theorem 2*

503 Assume that $\tilde{g}_{\tau_1}, \dots, \tilde{g}_{\tau_q} \in \mathcal{G}$ is the minimizing set of functions. Now consider any particular
 504 τ_t . The value of the loss function, the first term in (4), only depends on the values of g_{τ_t} at the data
 505 points. Hence any other g_{τ_t} such that $g_{\tau_t}(x) = \tilde{g}_{\tau_t}(x)$ at the data points will yield the same value
 506 for the loss function. So it suffices to consider the problem of finding the interpolator of the set of
 507 points $\{\tilde{g}_{\tau_t}(x_i)\}$ which minimizes the total variation. The solution to this interpolation problem
 508 is given by a linear spline with knots at the given set of x_i (Fisher and Jerome, 1975; Pinkus,
 509 1988). Denote this solution by \hat{g} . Clearly since the set of \tilde{g} are non-crossing, they maintain the
 510 proper ordering at each of the data points, hence the interpolating splines \hat{g} will not cross.
 511

512
 513 REFERENCES

514 Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2009). Improving point and interval
 515 estimators of monotone functions by rearrangement. *Biometrika* **96**, 559-575.
 516
 517 Cole, T. J. (1988). Fitting smoothed centile curves to reference data (with Discussion). *Journal*
 518 *of the Royal Statistical Society A* **151**, 385-418.
 519
 520 Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: The LMS method and
 521 penalized likelihood. *Statistics in Medicine* **11**, 1305-1319.
 522
 523 Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves.
 524 *Journal of the Royal Statistical Society B* **70**, 609-627.
 525
 526 Fisher, S. D. and Jerome, J. W. (1975). Spline solutions to L1 external problems in one and
 527 several variables. *Journal of Approximation Theory* **13**, 73-83.
 528
 529 Hall, P., Wolff, R. C. L., and Yao, Q. (1999). Methods for estimating a conditional distribution
 530 function. *Journal of the American Statistical Association* **94**, 154-163.

- 529 He, X. (1997). Quantile curves without crossing. *American Statistician* **51**, 186-192.
- 530 He, X. and Ng, P. (1999). COBS: Qualitatively constrained smoothing via linear programming.
531 *Computational Statistics* **14**, 315-337.
- 532
- 533 Jagger, T. H. and Elsner, J. B. (2009). Modeling tropical cyclone intensity with quantile regres-
534 sion. *International Journal of Climatology* **29**, 1351-1361.
- 535 Koenker, R. (1984). A note on L-estimators for linear models. *Statistics and Probability Letters*
536 **2**, 323-325.
- 537 Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.
- 538
- 539 Koenker, R. and Ng, P. (2003). SparseM: a sparse matrix package for R. *Journal of Statistical*
540 *Software* **8**, available at <http://www.jstatsoft.org/v08/i06>.
- 541 Koenker, R., Ng, P., and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika* **81**, 673-680.
- 542
- 543 Neocleous, T. and Portnoy, S. (2008). On monotonicity of regression quantile functions. *Statistics*
544 *and Probability Letters* **78**, 1226-1229.
- 545
- 546 Ng, P. and Maechler, M. (2007). A fast and efficient implementation of qualitatively constrained
547 quantile smoothing splines. *Statistical Modelling* **7**, 315-328.
- 548 Pinkus, A. (1988). On smoothest interpolants. *SIAM Journal of Mathematical Analysis* **19**, 1431-
549 1441.
- 550 Powell, J. L. (1991). Estimation of Monotonic Regression Models under Quantile Restrictions, in
551 *Nonparametric and Semiparametric Methods in Econometrics* (ed. by W. Barnett, J. Powell,
552 and G. Tauchen). Cambridge: Cambridge University Press.
- 553
- 554 R Development Core Team (2009). R: A language and environment for statistical comput-
555 ing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL
556 <http://www.R-project.org>.
- 557
- 558 Wu, Y. and Liu, Y. (2009). Stepwise multiple quantile regression estimation using non-crossing
559 constraints. *Statistics and Its Interface* **2**, 299-310.
- 560
- 561
- 562
- 563
- 564
- 565
- 566
- 567
- 568
- 569
- 570
- 571
- 572
- 573
- 574
- 575
- 576

577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624

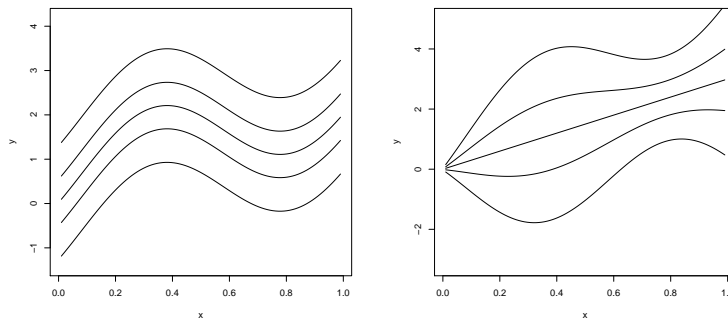


Fig. 1. Plot of the true conditional quantile functions for examples 4 (left) and 5 (right). The five curves from bottom to top represent $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$

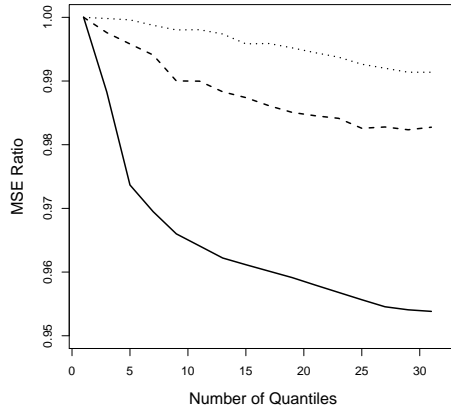


Fig. 2. Plot of Mean Squared Error in estimation of the slope at the median as a function of the number of included quantiles, for sample sizes $n = 50$ (solid line), $n = 100$ (dashed line), and $n = 200$ (dotted line). Each curve is scaled so that the Mean Squared Error is reported as a ratio relative to that of using only median regression.

Table 1. Average of root mean integrated squared error ($\times 100$) over 500 simulated data sets, with its standard error in parentheses.

	Example 1			Example 2		
	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$
NCRQ	30.1 (0.44)	40.7 (0.59)	72.9 (0.88)	42.9 (0.43)	53.2 (0.52)	89.7 (0.84)
RQ	31.2 (0.46)	42.9 (0.65)	86.1 (0.96)	47.9 (0.45)	66.5 (0.64)	121.3 (0.95)
RRQ	31.2 (0.46)	48.6 (0.70)	92.7 (2.01)	47.9 (0.45)	76.3 (0.70)	147.3 (2.34)
	Example 3, $n = 100$			Example 3, $n = 200$		
	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$
NCRQ	75.9 (0.92)	99.8 (1.19)	179.7 (2.04)	56.4 (0.66)	74.6 (0.91)	132.3 (1.62)
RQ	82.1 (0.99)	116.7 (1.36)	226.5 (1.85)	60.0 (0.69)	82.1 (0.99)	162.2 (1.74)
RRQ	82.1 (0.99)	133.0 (1.52)	250.7 (4.72)	60.0 (0.69)	91.5 (1.04)	158.9 (2.50)
	Example 3, $n = 500$					
	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$			
NCRQ	35.8 (0.41)	47.0 (0.55)	92.5 (1.14)			
RQ	37.1 (0.42)	49.7 (0.57)	109.0 (1.28)			
RRQ	37.1 (0.42)	56.7 (0.65)	96.0 (1.33)			

NCRQ, proposed non-crossing regression quantiles; RQ, classical regression quantiles; RRQ, restricted regression quantiles of He (1997).

625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672

673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720

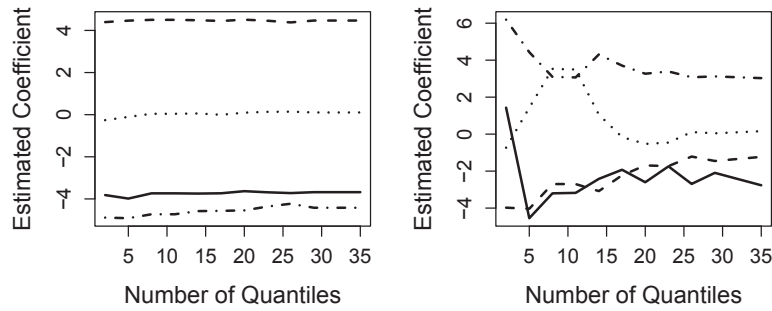


Fig. 3. Plot of the estimated slope coefficients at the median (left) and the .99 quantile (right) as the number of included quantiles is increased. NAO, North Atlantic Oscillation Index (solid line); SOI, Southern Oscillation Index (dashed line); SST, Atlantic sea surface temperature (dotted line); SUN, average sunspot number (dotted/dashed line).

Table 2. Average of root mean integrated squared error ($\times 100$) over 500 simulated data sets, with its standard error in parenthesis.

	Example 4			Example 5		
	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
NCRQ	25.7 (0.21)	25.9 (0.21)	31.8 (0.31)	26.4 (0.33)	32.2 (0.34)	48.5 (0.60)
NCRQ (single)	24.6 (0.19)	25.3 (0.19)	32.3 (0.34)	26.6 (0.34)	31.7 (0.32)	49.9 (0.62)
RQ (RS)	29.8 (0.18)	30.6 (0.21)	35.2 (0.31)	24.7 (0.42)	36.3 (0.43)	52.2 (0.73)
RQ (SS)	27.1 (0.19)	27.6 (0.21)	34.7 (0.31)	21.8 (0.35)	34.2 (0.39)	53.4 (0.90)

NCRQ, proposed non-crossing regression quantiles; NCRQ (single), proposed approach with a single tuning parameter; RQ (RS), classical regression splines with knot selection; RQ (SS), classical regression smoothing splines via regularization.

Table 3. *Coefficient estimates for the hurricane data at upper quantiles, with 90% confidence interval. The parameter estimates are given for the intercept and the four climate covariates.*

	Unconstrained			Constrained		
	$\tau = 0.9$	$\tau = 0.95$	$\tau = 0.99$	$\tau = 0.9$	$\tau = 0.95$	$\tau = 0.99$
INT	109.15 *	120.06 *	134.65 *	107.83 *	117.95 *	140.63 *
	(85.01, 133.29)	(88.26, 151.86)	(119.41, 149.89)	(84.56, 131.11)	(91.16, 144.74)	(121.44, 159.83)
NAO	-5.03 *	-0.95	1.43	-4.93 *	-3.06	-3.06
	(-9.91, -0.14)	(-6.38, 4.49)	(-1.58, 4.44)	(-9.61, -0.26)	(-8.04, 1.93)	(-6.54, 0.43)
SOI	5.74 *	1.23	-3.98 *	5.21 *	3.32	-2.80
	(0.40, 11.09)	(-5.75, 8.21)	(-7.44, -0.51)	(0.02, 10.39)	(-2.62, 9.27)	(-7.10, 1.50)
SST	6.16 *	4.19	-0.73	5.17 *	5.17 *	3.21
	(1.34, 10.97)	(-1.65, 10.03)	(-3.52, 2.06)	(0.57, 9.77)	(0.12, 10.22)	(-0.48, 6.90)
SUN	3.48	1.95	6.19 *	3.16	3.16	3.16
	(-1.21, 8.17)	(-2.60, 6.50)	(2.86, 9.52)	(-1.43, 7.75)	(-1.17, 7.49)	(-0.52, 6.84)

INT, intercept; NAO, North Atlantic Oscillation Index; SOI, Southern Oscillation Index; SST, Atlantic sea surface temperature; SUN, average sunspot number; *, statistically significant coefficients at $\alpha = 0.1$.