Joint Quantile Regression through Bayesian Semiparametrics

Surya T Tokdar
Duke University

Joseph B Kadane
Carnegie Mellon University

Abstract

We introduce a Bayesian semiparametric methodology for joint quantile regression with linearity and piecewise linearity constraints. We develop a probability model for all quantile curves in a continuum that define a coherent sampling distribution of the response variable. We provide a detailed illustration of model fitting and inference by analyzing wind speed trends of tropical cyclones in the North Atlantic.

Keywords: Bayesian inference, Gaussian process, Tropical cyclones, Change points, Markov chain Monte Carlo.

1 Introduction

In many real-world applications, interest focuses on how an explanatory variable $x \in \mathcal{X}$ affects the tails of the conditional distribution of a response $y \in \mathcal{Y}$. For example, in studying the effect of various demographic factors on birthweight (Abrevaya 2001) one might put special emphasis on the variations in the lower tail. Similarly, in analyzing the time trend of hurricane intensity (Elsner *et al.* 2008), one might pay extra attention to how the upper tail of the distribution of hurricane speeds changes over time. Such tail dependencies are impossible to capture through the traditional mean-regression models, such as the linear regression or the kernel regression. Quantile regression (Koenker and Bassett 1978), on the other hand, provides a natural platform for such analysis. In quantile regression, one models a given quantile of the conditional distribution as a function of the explanatory variable. An appropriate choice of the quantile point leads to a

direct inference on the tails. Unsurprisingly, the scientific literature on quantile regression has witnessed a steady growth in the last couple of decades, see, for example, Koenker (2005), Lancaster and Jun (2009) and Gelfand and Kottas (2003).

For almost all of the existing quantile regression methods, inference is done by specifying a separate model for each candidate value of the quantile. In some applications, a unique quantile of interest presents itself, e.g., the median for dealing with residual survival time (Gelfand and Kottas 2001). But in most cases, one can narrow down the choice of quantiles only to a well interpreted short range. Focusing on the bottom 10% tail of birth-weights, or the top 25% tail of hurricane speeds sounds reasonable from the point of view of a scientific investigation. But working only with the 0.1 quantile point or the 0.75 quantile point appears ad-hoc for these studies. In typical applications one chooses a number of representative points from the range of interest and presents the resulting analysis side by side (see Elsner *et al.* 2008 for an example). Such ensemble treatments, however, grossly overlook the fact that quantiles are closely intertwined in specifying a single distribution for the response. This allows the absurd possibility that the estimated quantiles can violate their natural ordering. More importantly, it is not at all clear how to combine these separate analysis to form a coherent single view of the effect of the explanatory variable on the response.

In this paper we introduce a novel framework for a joint regression analysis of all quantiles within a given range of interest. Our approach mirrors Bayesian density regression (see Dunson *et al.* 2007) where one specifies a family of conditional densities $\theta = \{f(y \mid x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ as the model parameter. Inference about functionals of θ is based on the posterior distribution $p(\theta \mid \text{data}) \propto \prod_{i=1}^{n} f(y_i \mid x_i)p(\theta)$ where $(x_i, y_i)_{i=1}^{n}$ are the observed data. Note that any quantile point of the conditional distribution can be expressed as a functional of θ (and x) – and hence density regression can be used for quantile regression as well. In this approach, however, the functional relationship between a quantile and x usually has a complicated nonparametric form. This makes this approach less appealing when one is mainly interested in inference about the trends of the quantile with respects to x.

To meet this goal, we directly specify as the model parameter a family of conditional quantiles $\theta = \{q(x, \tau), x \in \mathcal{X}, \tau \in [0, 1]\}$. The quantity $q(x, \tau)$ is to be interpreted as the number q for which $\Pr(y \leq q \mid x) = \tau$. This is uniquely defined whenever the condi-

tional distribution of y given x admits a density – a condition that we assume throughout the paper. For a coherent definition, we require for every $x \in \mathcal{X}$, $q(x, \tau_1) < q(x, \tau_2)$ whenever $\tau_1 < \tau_2$ with $q(x, 0) = y_{\min} := \inf \mathcal{Y}$ and $q(x, 1) = y_{\max} := \sup \mathcal{Y}$. An exact specification of θ that meets these requirements is discussed in the next section (see also Dunson and Taylor 2005 for an approximate Bayesian inference on a finite collection of quantiles based on a pseudo-likelihood function). The advantage of specifying θ in terms of $q(x, \tau)$ is that we can impose any desired structure on the functional form of the map $q_\tau(x) : x \mapsto q(x, \tau)$. For trend analysis, the most suitable form is a linear structure on $q_\tau(x)$ – and this is discussed in detail in Section 2. Section 3 illustrates an analysis of the tropical cyclone data (Elsner *et al.* 2008) with the proposed model. In Section 4, we discuss the case when a linear structure on $q_\tau(x)$ is imposed for only a specified range of τ values with a weak structure used for the rest. In Section 5 we relax the definition of $q_\tau(x)$ to a continuous, piecewise linear form. In these sections, we restrict our focus to a one-dimensional x for a clearer development of our nonparametric model. A simple multivariate extension is discussed in Section 6.

2 A Joint Model for Linear Quantile Regression

We begin with the case where linearity is desired for $q_\tau(x)$ for all $\tau \in [0, 1]$. We shall restrict our focus to $\mathcal{X}$ given by a compact subset of an Euclidean space. For clarity of exposition we shall assume x to be one-dimensional and take $\mathcal{X} = [0, 1]$ by applying a suitable location and scale change to the original explanatory variable. We further assume $\mathcal{Y}$ to be bounded as well – and transform it to $[0, 1]$. This assumption is not binding for most real-world data, it is perhaps quite reasonable to state that birthweight of human babies cannot exceed 20lb, or that the maximum wind speed of a North Atlantic hurricane must be under 200 nautical miles (nm) per hour.

Under these assumptions, $q(x, \tau)$ must satisfy

$$q(x, \tau) = (1 - x)q(0, \tau) + xq(1, \tau) \tag{1}$$

with $q(0, \tau)$ and $q(1, \tau)$ defining monotonically increasing functions in τ with $q(0, 0) = q(1, 0) = 0$ and $q(0, 1) = q(1, 1) = 1$. In other words $q(0, \tau)$ and $q(1, \tau)$ define two cumulative distribution functions (CDF)

on $\tau \in [0,1]$. For ease of discussion, we introduce the conditional quantile functions $Q_x : \tau \mapsto q(x,\tau)$ and re-express (1) as

$$Q_x = (1-x)Q_0 + xQ_1. \tag{2}$$

In fact, with Q_0 and Q_1 as CDFs on $[0,1]$, (2) gives a complete characterization of $\theta = \{q(x,\tau), x \in [0,1], \tau \in [0,1]\}$ with linearity constraints on q_τ. Because we assumed that the conditional distribution of y given x admits a density $f(y \mid x)$, Q_0 and Q_1 must also be differentiable and satisfy

$$f(y \mid x) = \frac{1}{Q'_x(Q_x^{-1}(y)}. \tag{3}$$

Thus a Bayesian analysis of θ given observations (x_i, y_i), $1 \le i \le n$ can be based on the posterior distribution

$$p(Q_0, Q_1 \mid \text{data}) \propto \left[\prod_{j=1}^{n} \frac{1}{Q'_{x_j}(Q_{x_j}^{-1}(y_j))}\right] p(Q_0, Q_1)$$

by specifying a suitable prior distribution $p(Q_0, Q_1)$ on Q_0 and Q_1.

To specify $p(Q_0, Q_1)$, we consider a random family of CDFs on $[0,1]$, $\Phi = \{\phi_x(\cdot) \mid x \in [0,1]\}$ for which $x \mapsto \phi_x$ is smooth (almost surely) and equate $Q_0(\tau) = \phi_0(\tau)$ and $Q_1(\tau) = \phi_1(\tau)$. Such an embedding allows us to model the covariance between Q_0 and Q_1 – a quantity that directly relates to the slopes of the lines $q_\tau(x)$. In actual model fitting, however, only the extreme members ϕ_0 and ϕ_1 of Φ play a role. We define Φ through an appropriate transformation of a smooth sieve-Gaussian process (sGP) as discussed below.

For a real valued smooth function $\eta(x,\tau)$ on $x \in [0,1]$ and $\tau \in [0,1]$

$$\phi_x(\tau) = \frac{\int_0^\tau e^{\eta(x,t)} dt}{\int_0^1 e^{\eta(x,t)} dt}, x \in [0,1], \tau \in [0,1]$$

defines a CDF ϕ_x on $[0,1]$ that varies smoothly with x. Therefore a random Φ can be constructed simply by taking η to be a smooth Gaussian process (GP) on $[0,1] \times [0,1]$. This choice, however, leads to an intractable posterior computation. A more useful choice is given by

$$\eta(x,\tau) = \mathbb{E}[\omega(x,\tau) \mid \omega(x_1^*, \tau_1^*), \cdots, \omega(x_K^*, \tau_K^*)]$$

where ω is a smooth GP on $[0,1]\times[0,1]$ and $\mathcal{N} = \{(x_1^*, \tau_1^*), \cdots, (x_K^*, \tau_K^*)\}$ is a finite set of points from $[0,1] \times [0,1]$. The process $\eta(x,\tau)$, which

we call a sieve-Gaussian Process (sGP), provides an interpolation type approximation to ω but is completely determined through the K dimensional random vector $\omega_{\mathcal{N}} = (\omega(x_1^*, \tau_1^*), \cdots, \omega(x_K^*, \tau_K^*))'$.

We take ω to be a mean-zero GP with its covariance function given by the square-exponential kernel $\mathbb{C}\text{ov}(\omega(x_1, \tau_1), \omega(x_2, \tau_2)) = \sigma^2 \exp[-\beta_x^2(x_1 - x_2)^2 - \beta_\tau^2(\tau_1 - \tau_2)^2]$. The corresponding $\eta(x, \tau)$ can be written as

$$\eta(x, \tau) = \sum_{k=1}^{K} \xi_k \exp[-\beta_x^2(x - x_k^*)^2 - \beta_\tau^2(\tau - \tau_k^*)^2]$$

where $\xi = (\xi_1, \cdots, \xi_K)' = \Sigma^{-1}\omega_{\mathcal{N}} \sim \text{Normal}(0, \Sigma^{-1})$ with the (k, l)-th of element of Σ given by $\sigma^2 \exp[-\beta_x^2(x_k^* - x_l^*)^2 - \beta_\tau^2(\tau_k^* - \tau_l^*)^2]$. We fix $\mathcal{N}$ as

$$\mathcal{N} = \{(0, 0), (0, \delta), (0, 2\delta), \cdots, (0, 1), (1, 0), (1, \delta), (1, 2\delta), \cdots, (1, 1)\}$$

for some small δ (such as $\delta = 0.1$). This choice places two equi-spaced grids on the (x, τ)-space, one for $x = 0$ and the other for $x = 1$, each covering the entire range of τ. This is a reasonable choice since our model for θ is based only on $\eta(0, \tau)$ and $\eta(1, \tau)$. In Section 5 we discuss the case where $\mathcal{N}$ is modeled with a uniform distribution.

The parameters β_x, β_τ and σ^2 all exert control over the variation of $\eta(x, \tau)$. Among these, β_x has the biggest impact on the slopes $Q_1(\tau) - Q_0(\tau)$ of the regression lines $q_\tau(x)$. For the limiting case $\beta_x = 0$, the two curves Q_0 and Q_1 are identical, and when $\beta_x \to \infty$, these become independent of each other. Similarly, β_τ, which affects the covariance between $\eta(\tau_1, x)$ and $\eta(\tau_2, x)$, determines how each of Q_0 and Q_1 increases from 0 at $\tau = 0$ to 1 at $\tau = 1$. A small value of β_τ ensures that this increase is close to linear in τ and hence the conditional density of y given x is approximately uniform. A very large value of β_τ allows Q_0 and Q_1 to increase through a series of short bursts, making the conditional densities possibly multimodal. A range of different types of increase are entertained for values in between. The overall variation of $\eta(x, \tau)$, and consequently of Q_0, Q_1 as well as $Q_1 - Q_0$, is controlled by σ^2.

We use thin-tail priors on β_τ and β_x and a rather diffuse inverse-chisquare prior on σ^2:

$$p(\beta_x, \beta_\tau, \sigma^2) = \text{ExGam}(\beta_x \mid \nu_x, \mu_x) \times \text{ExGam}(\beta_\tau \mid \nu_\tau, \mu_\tau) \times \text{Inv-}\chi^2(\sigma^2 \mid \nu_\sigma, \sigma_0^2).$$

Here $\beta \sim \text{ExGam}(\nu, \mu)$ – the extreme-gamma distribution with shape ν and scale μ – means that $\beta = \mu \log(1 + \gamma)$ where γ follows a gamma

distribution with shape ν and unit scale. And $\sigma^2 \sim \text{Inv-}\chi^2(\nu, \sigma_0^2)$ means that $\kappa = 1/\sigma^2$ follows a gamma distribution with shape $\nu/2$ and scale $2/(\nu\sigma_0^2)$. Notice that the right tail of $\text{ExGam}(\nu, \mu)$ decreases extremely fast. If $\beta \sim \text{ExGam}(\nu, 1)$, then $\Pr(\beta > b) = O(\exp(-\exp(b)))$.

3 Illustration: Trends of Tropical Cyclones

Elsner *et al.* (2008) argued that tropical cyclones (TCs) in the North Atlantic basin are getting stronger over the years. Figure 1 shows a scatter plot of the maximum wind speed (`WmaxST`, derived from satellite images) of all North Atlantic tropical cyclones between 1981 and 2006 against their year of occurrence (`Year`). We analyzed this data with our proposed joint linear quantile regression model on `WmaxST` with `Year` as the explanatory variable. We took `WmaxST` to be bounded between 0 and 200, whereas the range of interest for the explanatory variable was taken to be $[1981, 2006]$. These bounds were used to transform these two variables into the interval $[0, 1]$. However, all results presented here are shown in the original scale.

We fixed the prior parameters at: $\nu_0 = 3$, $\sigma_0^2 = 1$, $\nu_\tau = 2$, $\mu_\tau = 5$, $\nu_x = 2$ and $\mu_x = 2$. The corresponding prior distribution on the slope s_τ of the quantile line q_τ is summarized in Figure 2 over a grid of τ values in $[0, 1]$. The top left panel of this figure shows the point-wise medians (black beads connected by a line), equal-tail 50% intervals (mesh of dark grey vertical lines) and equal-tail 95% intervals (mesh of light grey vertical lines in the background) for all τ in $\{0, 0.01, \cdots, 1\}$. The same is repeated in the bottom left panel but only for τ values larger than or equal to 0.75 – the range of interest for this analysis. Other choices of the prior parameters led to similar posterior inference and are not reported here.

For posterior inference, we used a Markov chain sampler to draw $(\omega_{\mathcal{N}}, \beta_x^2, \beta_\tau^2)$ from their joint posterior distribution given the data. The variance parameter σ^2 was integrated out by using the conjugacy property of the normal distribution on $\omega_{\mathcal{N}}$ and the inverse chi-square distribution on σ^2. Each coordinate of $\omega_{\mathcal{N}}$ was updated through a symmetric Metropolis move generated from a t distribution, whereas β_x^2 and β_τ^2 were updated through asymmetric Metropolis-Hastings moves generated from log-normal distributions. The likelihood com-

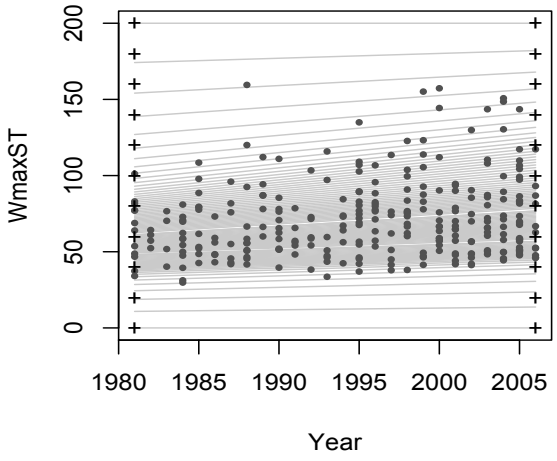

Figure 1: Tropical cyclone data form North Atlantic basin (Elsner *et al.* 2008). Scatter plot of maximum wind speed (in nautical miles per hour) versus year of occurrence. The light grey lines in the background show the quantile lines $q_\tau(x)$ from one iteration of our MCMC for fitting the joint linear quantile regression model.

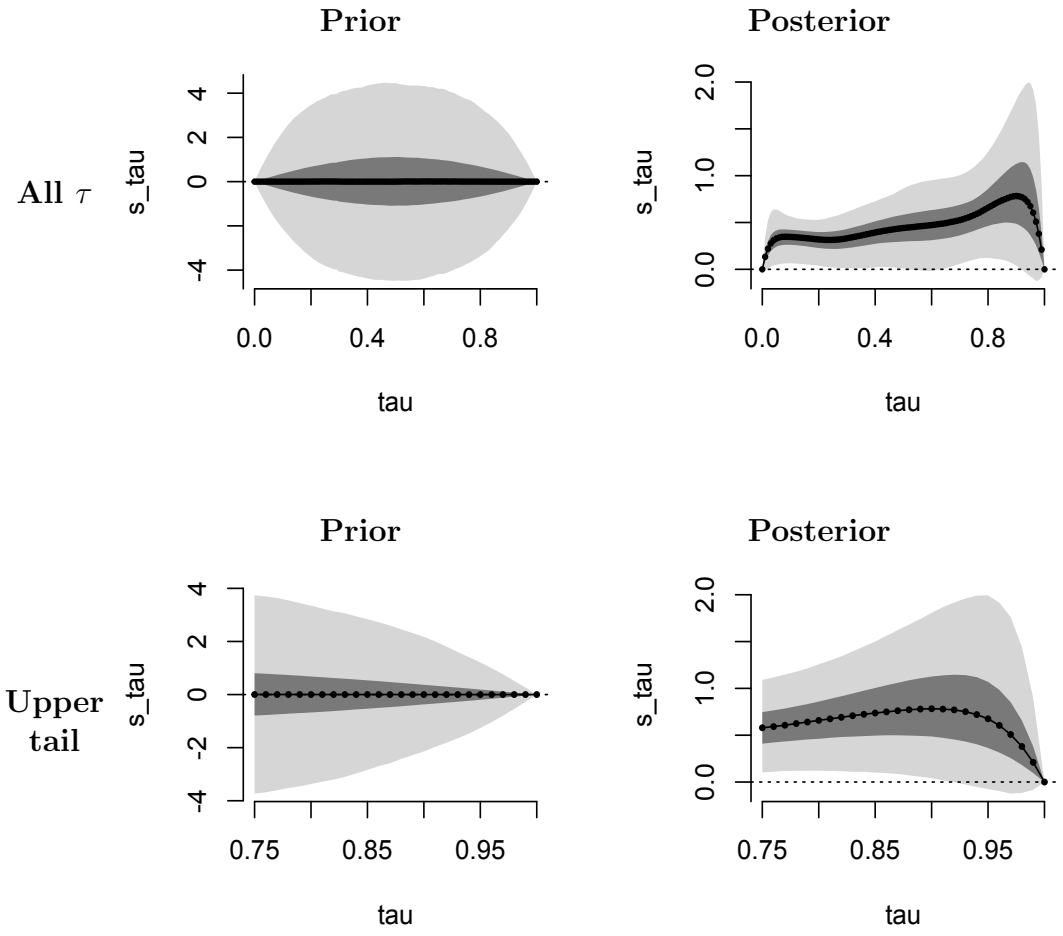

Figure 2: Prior and posterior credible intervals for slopes of quantile regression lines for the analysis of Tropical Cyclone data from Elsner *et al.* (2008). The top row shows slopes for a grid of τ covering the entire interval $[0, 1]$ with mesh size 0.01. The bottom row shows an enlargement of the upper tail: $\tau \in [0.85, 1]$. Note the drastic rescaling of the y-axis between the prior (left) and the posterior (right).

putation was carried out by evaluating $\eta(0,\tau)$ and $\eta(1,\tau)$ at $\tau \in \{0.00, 0.01, 0.02, \cdots, 1.00\}$ and approximating $\int_0^\tau e^\eta(x,t)dt$ on this grid points by the trapezoid rule. The resulting discretized version of Q_x was then used to approximate the density evaluations $f(y_j \mid x_j) = 1/Q'_{x_j}[Q^{-1}_{x_j}(y_j)]$.

We ran four separate chains of the sampler, over-dispersed with respect to the initial value of $\omega_{\mathcal{N}}$. This ensured over-dispersion with respect to the slopes s_τ. Each chain was run for 100,000 iterations. The chains converged rapidly with respect to s_τ (see Figure 3) and other parameters, as per the *potential scale reduction factor* check of Gelman and Rubin (1992). The first 10,000 iterations were removed from each chain, and the remaining iterations were thinned and pooled together to give a sample of size 10,000.

The grey lines in the background of Figure 1 represent the actual q_τ lines derived from a single draw in our Markov chain sample (MCS). The top and bottom right panels of Figure 2 show pointwise medians and equal-tail 50% and 95% intervals of the posterior distribution of s_τ, as approximated by our MCS. Table 1 reports the approximated posterior probabilities of the events $\{s_\tau > 0\}$ for $\tau \in \{0.85, 0.9, 0.95, 0.975, 0.99\}$.

Our summary in Table 1 supports the claim by Elsner *et al.* (2008) that the strongest North Atlantic TCs have gotten stronger over the last couple of decades. However, we notice a greater cohesion between the posterior probabilities in Table 1 as compared to the P-values reported in Table 1 of Elsner *et al.* (2008) where it appears that while s_τ is significantly different from zero for $\tau = 0.9$ and $\tau = 0.975$, it is not for $\tau = 0.95$!

A closer look at Figure 2 reveals that an increasing trend persists across the whole range of τ – a finding that differs from the conclusions drawn by Elsner *et al.* (2008). However, the magnitude of this increase is more dramatic in the upper tail. These findings have potential implications for the current debate on global warming (Trenberth 2005).

4 Linear model on a subinterval of τ

Next we consider the case when a joint linear model for q_τ is desired only for τ within a subinterval of $[0,1]$. For ease of exposition we fix this subinterval to be $[\underline{\tau}, 1]$ for some $0 < \underline{\tau} < 1$. We pursue a simple

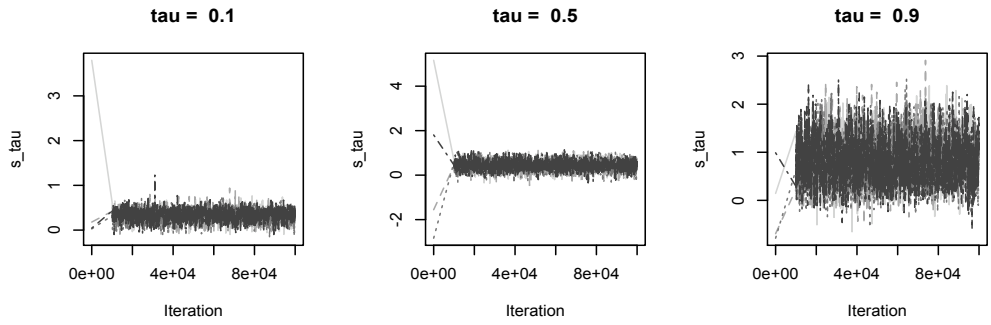

Figure 3: Trace plots of s_τ for $\tau = 0.1, 0.5, 0.9$ from four parallel, thinned out chains.

North Atlantic Basin					
τ	0.85	0.90	0.95	0.975	0.99
Pr(slope of $q_\tau > 0 \mid$ data)	0.989	0.980	0.960	0.942	0.921

Table 1: Posterior probability that a quantile line q_τ has a positive slope, for chosen values of τ from the upper tail of `WmaxST`.

extension of the model described in Section 2 and specify a very weak structure on the conditional behavior of y_j given $y_j < q_{\underline{\tau}}(x_j)$. First, a linear model is specified on the boundary line $q_{\underline{\tau}}$ through

$$q_{\underline{\tau}}(x) = (1-x)\underline{Q}_0 + x\underline{Q}_1$$

where $(\underline{Q}_0, \underline{Q}_1) \in (0,1)^2$. Next the quantile lines $q_\tau(x) = Q_x(\tau)$ for $\tau > \underline{\tau}$ are defined as $Q_x(\tau) = (1-x)Q_0(\tau) + xQ_1(\tau)$ with

$$Q_\ell(\tau) = \underline{Q}_\ell + (1-\underline{Q}_\ell)\phi_\ell\left(\frac{\tau - \underline{\tau}}{1-\underline{\tau}}\right), \quad \ell = 0, 1$$

where ϕ_x are as in Section 2. Then the conditional density $f(y_j \mid x_j)$ at an observation point $1 \le j \le n$ can be split as

$$f(y_j \mid x_j) = \underline{\tau} f_j(y_j \mid x_j) + I(y_j \ge q_{\underline{\tau}}(x_j)) \frac{1}{Q'_{x_j}(Q_{x_j}^{-1}(y_j))}$$

with $f_j(y \mid x_j)$ defining a probability density function with support $[0, q_{\underline{\tau}}(x_j)]$.

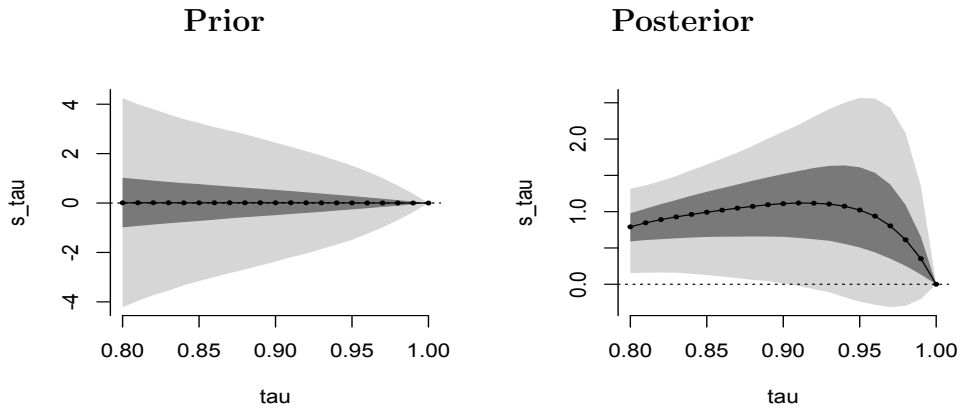

Figure 4: Prior and posterior credible intervals for slopes of quantile regression lines for the analysis of Tropical Cyclone data from Elsner *et al.* (2008). Only q_τ for $\tau > 0.8$ were modeled.

We take $\underline{Q}_\ell = 1/(1+\exp(-\lambda_\ell))$, $\ell = 0, 1$ where (λ_0, λ_1) is taken to be a bivariate normal vector, independent of η, with each component having mean $\text{logit}(\underline{\tau})$ and variance σ^2_λ and their correlation given by $\exp(-\beta_x^2)$. We specify prior distributions on β_x, β_τ and σ^2 as in Section 2, possibly with different hyper-parameters, and set σ^2_λ fixed. We model f_j's, $j = 1, \cdots, n$ as conditionally independent given η, q_0, q_1 with $\mathbb{E}[f_j(y \mid x_j)] = \text{Uniform}(y \mid 0, q_{\underline{\tau}}(x_j))$. This assumption of conditional independence across j allows us to marginalize over f_j's producing the integrated likelihood of $(\eta, \underline{Q}_0, \underline{Q}_1)$ as

$$\prod_{j=1}^{n} \left[\left(\frac{\underline{\tau}}{q_{\underline{\tau}}(x_j)} \right)^{I(y_j < q_{\underline{\tau}}(x_j))} \left(\frac{1}{Q'_{x_j}(Q^{-1}_{x_j}(y_j))} \right)^{I(y_j \geq q_{\underline{\tau}}(x_j))} \right].$$

Therefore, for posterior inference on q_τ, $\tau \geq \underline{\tau}$, we do not require to explicitly model f_j except for specifying its mean. As a consequence, the posterior computation remains very similar to the one required for the model in Section 2. One must remember, however, that this model offers no learning and prediction for the conditional quantiles of y at $\tau < \underline{\tau}$.

Figure 4 shows a graphical summary of an analysis of the North Atlantic tropical cyclone data with the new model with $\underline{\tau} = 0.8$. As before, likelihood computation was carried out on a grid of τ values spanning $(\underline{\tau}, 1)$ with an increment of 0.01. The hyper-parameters on

β_x, β_τ and σ^2 were chosen as in Section 3 and σ_λ^2 was fixed at 2. The left panel of Figure 4 shows the median and 50% and 95% credible intervals from the prior distribution of s_τ, the slope of q_τ on a grid of τ values. The right panel shows the same summaries from the posterior distribution of s_τ. We approximated the posterior intervals through a Markov chain sampler as in Section 3, only with an additional step per iteration for a symmetric Metropolis update of λ. We ran four parallel chains, each for 100,000 iterations of which the first 10,000 iterations were discarded as burn-in, and the remaining 90,000 iterations were evenly thinned to a total of 2,500 draws per chain. The chains converged rapidly – they were initialized with over-dispersed starting values for s_τ. The thinned chains passed the potential scale reduction factor check for convergence. They were then combined to form a Markov chain sample of size 10,000.

The posterior summary of s_τ, $\tau > 0.8$, under the new model is quite similar to the one obtained in Section 3, the only difference being a slight upward movement and widening of the credible intervals under the new model. The posterior probability of $\{s_\tau > 0\}$ for $\tau \in \{0.85, 0.9, 0.95, 0.975, 0.99\}$ were approximated to be 0.990, 0.976, 0.942, 0.919 and 0.892, respectively.

5 Piecewise linear joint quantile regression

For trend analysis we argue in Section 2 that linear quantile lines are much more interpretable than arbitrary smooth curves derived from a non-parameteric density regression model. It is however possible to capture much of the flexibility of the latter by constructing a piecewise linear model on the quantiles that retains some of the interpretability of a simple linear model. In this section we discuss a straightforward extension of our joint linear model of Section 2 to a piecewise linear one. The same can be applied to the restricted model of Section 4, but we omit the details.

The basic premise of our extension is that the same set of change points apply to all quantile lines in determining their piecewise linearity. That is, $\mathcal{X} = [0, 1]$ can be split into to subintervals $[b_{m-1}, b_m]$ with $0 = b_0 < b_1 < \cdots < b_M = 1$, so that within each subinterval $q_\tau(x)$ are linear in x for all τ. This essentially postulates possible global regime changes in the conditional behavior of the response, but within any

regime a single linear model on the quantiles suffices to explain the dependence of y on x. We also restrict these changes to be continuity preserving, so that each $q_\tau(x)$ remains continuous over the whole of $x \in [0,1]$. Such a model on $q_\tau(x) = Q_x(\tau)$ can be constructed as

$$Q_x(\tau) = \frac{(b_m - x)Q_{b_{m-1}}(\tau) + (x - b_{m-1})Q_{b_m}(\tau)}{b_m - b_{m-1}} \text{ for } x \in (b_{m-1}, b_m],$$

where Q_{b_m} defines a CDF on $\tau \in [0,1]$ for $m = 0, 1, \cdots, M$. Now consider $\Phi = \{\phi(x,\tau)\}$ – a random family of CDFs on $\tau \in [0,1]$ which vary smoothly with respect to $x \in [0,1]$. Then we can put a joint model on Q_{b_m}, $0 \le m \le M$ by setting $Q_{b_m}(\tau) = \phi(b_m, \tau)$. Our choice of Φ from Section 2, namely $\phi(x,\tau) = \int_0^\tau \exp(\eta(x,t))dt / \int_0^1 \exp(\eta(x,t))dt$ where $\eta(x,\tau)$ is a sieve-GP, fits in perfectly with this framework. The node set $\mathcal{N}$ however requires an expansion to the whole of $\mathcal{X} = [0,1]$, and we shall do this by considering the elements (x_k^*, τ_k^*) of $\mathcal{N}$ as model parameters and take them to be uniformly distributed over $[0,1] \times [0,1]$.

To specify a prior on the change points, first define the interval lengths $\gamma_m = b_m - b_{m-1}$, $m = 1, \cdots, M$. The vector $\gamma = (\gamma_1, \cdots, \gamma_M)$ defines a probability vector of length M. One can thus proceed by defining a prior $p(M)$ on M and then model $p(\gamma \mid M) = \text{Dir}(\gamma \mid \alpha_{M,1}, \cdots, \alpha_{M,M})$. We arrive at such a model through a slightly roundabout way in order to facilitate computation as well as the choice of the parameters $\alpha_{M,m}$. We first choose an upper bound $M_{\max}$ on M. Next introduce $\lambda_l \in \{1, \cdots, M_{\max}\}$, $1 \le l \le M_{\max}$ and $\chi = (\chi_1, \cdots, \chi_{M_{\max}})$ – a probability vector – and define

$$\gamma_m = \sum_l I(\lambda_l = m)\chi_l, \text{ for } m = 1, \cdots, M_{\max}.$$

This defines $M_{\max}$ disjoint intervals $(b_{m-1}, b_m]$ covering $(0,1]$ where $b_m = \gamma_1 + \cdots + \gamma_m$. Some of these intervals, however, could be empty – $(b_{m-1}, b_m]$ has length zero whenever there is no λ_l that equals m. Therefore the number of non-empty intervals defined as above can vary from $M = 1$ to $M = M_{\max}$ depending on the ties within the vector $\lambda = (\lambda_1, \cdots, \lambda_{M_{\max}})$.

We induce ties within λ by specifying

$$p(\lambda) = \frac{c^{k-1} \prod_{l=1}^k (N_l - 1)!}{\prod_{l=1}^{M_{\max}} (c + l)}$$

where k denotes the number of distinct elements in λ and $N_1, \cdots, N_k$ denote the sizes of the k clusters formed by partitioning λ according to its ties, and c is a fixed positive number. This formulation is

exactly similar to the clustering distribution induced by a Dirichlet process prior (see Escobar and West 1995). We model χ as $p(\chi) = \text{Dir}(\chi|a, \cdots, a)$ where $a > 0$ is a fixed number.

The above specification has two advantages. First the conditional distribution of interval lengths can be written as $p(\gamma \mid \lambda) = \text{Dir}(aK_1, \cdots, aK_{M_{\max}})$ where $K_m = \#\{l : \lambda_l = m\}$. This allows a greater flexibility in how the intervals can vary from each other in length, although, marginalized over λ, they remain exchangeable. Second, the models corresponding to two consecutive choices of M, say $M = m$ and $M = m + 1$, are not entirely isolated. One can move from one model to another through a series of small changes in the ties of λ.

This new model can be fitted with the help of a Markov chain sampler similar to the one described in Section 2, but with additional updates for λ and χ. To update χ given λ and other parameters, we reparametrize the model in terms of $\gamma_m = \sum_l I(\lambda_l = m)\chi_l$ and $\chi^*_{(m)} = \text{vector}(\{\chi_l : \lambda_l = m\})/\gamma_m$, $1 \le m \le M_{\max}$ where vector(S) denotes the vector formed by the elements of a set S (placed in any arbitrary order). The conditional prior distribution on $(\gamma = (\gamma_1, \cdots, \gamma_{M_{\max}}), \chi^*_{(1)}, \cdots, \chi^*_{(M_{\max})})$, given λ and other parameters, is

$$\text{Dir}(\gamma \mid n_1 a, n_2 a, \cdots, n_{M_{\max}} a) \prod_{m=1}^{M_{\max}} \text{Dir}(\chi^*_{(m)} \mid a, \cdots, a)$$

where $n_m = \sum_{l=1}^{M_{\max}} I(\lambda_l = m)$. We update γ through a Metropolis-Hastings move and update $\chi^*_{(m)}$'s by sampling these vectors from their conditional posterior distributions which are simply identical to their conditional prior distributions described above. To update λ_m, we propose a new value for it by sampling from the conditional prior $p(\lambda_m \mid \{\lambda_l : l \ne m\})$ and then accept or reject it according to the Metropolis-Hastings acceptance formula. This conditional prior distribution on λ_m can be represented by a probability vector $\pi^{(m)} = (\pi^{(m)}_1, \cdots, \pi^{(m)}_{M_{\max}})$ on $\{1, \cdots, M_{\max}\}$ with $\pi^{(m)}_l \propto n^-_l$ when $n^-_l = \sum_{j \ne m} I(\lambda_j = l) > 0$ and $\pi^{(m)}_l \propto c/(M_{\max} - k)$ otherwise, where $k = \sum_{l=1}^{M_{\max}} I(n^-_l > 0)$.

Figure 5 shows a summary of our analysis of the cyclone data with the proposed piecewise linear model. The top panel shows snapshots of q_τ curves from two iterations of our MCMC. We used the same ν_0, σ^2_0, ν_τ and μ_τ as in Section 3 but fixed $\nu_x = 5$ and $\mu_x = 5$. This change allows the prior distribution on β_x to shift to the right, entertaining

Prior probabilities for piecewise linear model					
# Subintervals	1	2	3	4	5
Prior probability	0.039	0.147	0.249	0.253	0.175
# Subintervals	6	7	8	9	10
Prior probability	0.089	0.034	0.010	0.003	0.000

Table 2: Prior probability distribution of number of subintervals

more variability in $\phi(x, \cdot)$ across x in anticipation of breaks in linearity. We took $M_{\max} = 25$ – the total duration (in years) of the observation period 1981-2006. We fixed the precision constant $c = 1$, the induced prior distribution on the number of subintervals is reported in Table 2. We used $a = 0.5$, allowing the marginal density of each χ_l to peak at zero. Such a choice allows more rapid trades of λ_l between different subintervals, since each such change creates a small perturbation in the likelihood function.

The left bottom panel of Figure 5 shows the posterior distribution of the number of subintervals. This summary alone may not suffice to make inference on the actual breaks in linearity. Our model is prone to maintaining short subintervals that move rapidly in order to assist bridging configurations with different number of change points. For this reason, we provide another summary of the change points in the right bottom panel of Figure 5. Here we show the posterior probability (color coded in grey scale) that each of the time bins 1981-1982, 1982-1983 and so on, contains a boundary of at least one non-empty subinterval. It appears that apart from the two boundary bins (1981-1982 and 2005-2006) the other bins have negligible posterior probabilities (≤ 0.10) of containing any change point. The boundary bins (0.38 and 0.47 respectively) have moderate probabilities of containing one. Inference at boundaries, however, should always be done with more caution!

6 Multivariate extension

In Sections 2 and 4, we discussed our joint linear quantile regression models only in the context of a univariate x. These models can be easily extended to the multivariate case, but some additional computations are required. We shall describe this only for the model from Section 2, where a linear model is sought for all $\tau \in [0, 1]$. When

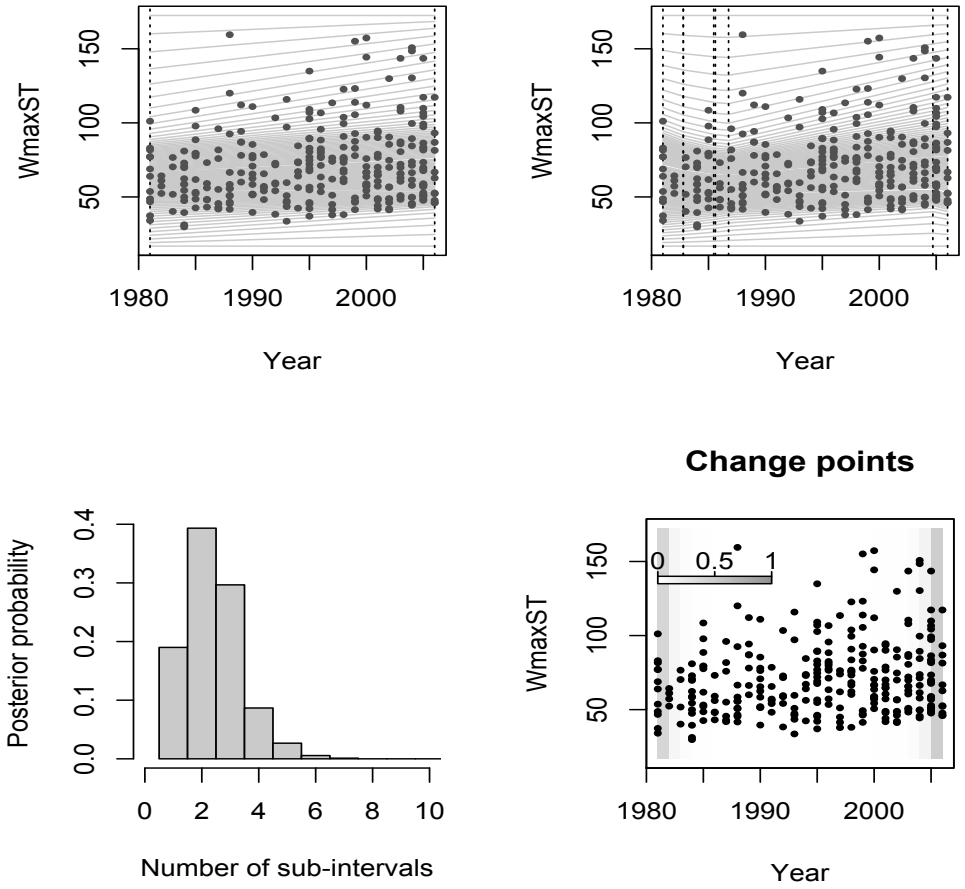

Figure 5: Analysis of cyclone data with piecewise linear joint quantile regression model. The top row shows snapshots of q_τ curves from two separate iterations of our MCMC. The bottom row shows two summaries of the subinterval formation. The left plot shows the posterior distribution on the number of subintervals. The right plot shows the posterior probability for each time bin (1981-1982), (1982-1983) and so on, to contain at least one change point. These posterior probabilities are color coded in a grey scale with darker bins having a higher posterior probability of containing a change point. No change points are apparent except for the two bins at the boundary.

$x = (x_1, \cdots, x_d) \in [0,1]^d$, one needs $d+1$ many CDFs, Q_0, Q_1, $\cdots$, Q_d, to define the quantile curves $q_\tau(x)$ as

$$q_\tau(x) = (1 - x_1 - x_2 - \cdots - x_d)Q_0(\tau) + x_1 Q_1(\tau) + \cdots + x_d Q_d(\tau).$$

However, for $d > 1$, such a definition would not guarantee that $q_{\tau_1}(x) < q_{\tau_2}(x)$ for all $x \in [0,1]^d$ whenever $\tau_1 < \tau_2$. Therefore, one needs to restrict the model $Q_0, \cdots, Q_d$ to the set for which q_τ defined as above obeys these quantile orderings. One can still use the sieve-Gaussian process construction: $Q_j(\tau) = \phi(e_j, \tau)$ with $\phi(x,\tau) = \int_0^\tau \exp(\eta(x,t))dt / \int_0^1 \exp(\eta(x,t))dt$, where $\eta(x,\tau)$ defines a sieve-GP on $[0,1]^d \times [0,1]$ and e_j denotes the j-th canonical vector in $\mathbb{R}^d$ with all zeros except for a one at its j-th coordinate. The Markov chain exploration of this model now must begin with an η for which q_τ is correctly ordered at every $x \in [0,1]^d$, and the subsequent updates must reject all proposals that would lead to violation of this property. The ordering checks are easy to perform once we restrict ourselves to a grid of τ values $\{0, \delta, 2\delta, \cdots, 1\}$ finely spanning $[0,1]$. It then suffices to check for every pair of values $\tau_1 = m\delta$ and $\tau_2 = (m+1)\delta$ that $q_{\tau_1}(x) < q_{\tau_2}(x)$ at every $x \in [0,1]^d$. Because q_τ's are hyper-planes, $q_{\tau_1}(x) < q_{\tau_2}(x)$ for all $x \in [0,1]^d$ if and only if $q_{\tau_1}(\tilde{x}) < q_{\tau_2}(\tilde{x})$ where $\tilde{x} = (\tilde{x}_1, \cdots, \tilde{x}_d)$ is given by

$$\tilde{x}_j = I(Q_j(\tau_1) - Q_j(\tau_2) > Q_0(\tau_1) - Q_0(\tau_2)), \;\; j = 1, \cdots, d.$$

Therefore it suffices to check the ordering of q_{τ_1} and q_{τ_2} only at the point $\tilde{x}$.

7 Conclusion

We have introduced a statistical framework for a joint analysis of quantile regression on a range of quantile points with linearity and piecewise linearity constraints on the quantile curves. This framework offers a sounder alternative to the usual practice of stitching together quantile regression analysis done separately for each choice of a quantile point from a range of interest. Our analyses of the North Atlantic tropical cyclone data shows interesting differences with a previous analysis of this data by Elsner *et al.* (2008). We find that an increasing trend of intensity persists across the entire range of quantiles and the rate of this increase heightens in the upper tail.

Acknowledgements

We thank Anne-Sophie Charest for her help with the data set and comments on the manuscript.

References

Abrevaya, J. (2001). The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics*, Springer, **26**, 247-257.

Dunson, D. B., Pillai, N. S. and Park, J-H. (2007). Bayesian density regression. *Journal of the Royal Statistical Society B*, **69**, 163-183.

Dunson, D. B. and Taylor, J. A. (2005). Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics*, **17**, 385-400.

Elsner, J. B., Kossin, J. P. and Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature*, **455**, 92-95.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 577-588.

Gelfand, A. E. and Kottas, A. (2003). Bayesian Semiparametric Regression for Median Residual Life'. *Scandinavian Journal of Statistics*, **30**, 651-665.

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, **7**, 457-472.

Koenker, R. (2005). Quantile Regression. Cambridge U. Press.

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

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric mean regression modeling. *Journal of the American Statistical Association*, **96**, 1458-1468.

Lancaster, T. and Jun, S. J. (2009). Bayesian Quantile Regression Methods. *Journal of Applied Econometrics*, forthcoming.

Trenberth, K. (2005). Uncertainty in Hurricanes and Global Warming. *Science*, **308**,1753-1754.