Bayesian Analysis (2011) **6**, Number 4, pp. 1–22

Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach

Surya T. Tokdar* and Joseph B. Kadane†

Abstract. We introduce a semi-parametric Bayesian framework for a simultaneous analysis of linear quantile regression models. A simultaneous analysis is essential to attain the true potential of the quantile regression framework, but is computationally challenging due to the associated monotonicity constraint on the quantile curves. For a univariate covariate, we present a simpler equivalent characterization of the monotonicity constraint through an interpolation of two monotone curves. The resulting formulation leads to a tractable likelihood function and is embedded within a Bayesian framework where the two monotone curves are modeled via logistic transformations of a smooth Gaussian process. A multivariate extension is suggested by combining the full support univariate model with a linear projection of the predictors. The resulting single-index model remains easy to fit and provides substantial and measurable improvement over the first order linear heteroscedastic model. Two illustrative applications of the proposed method are provided.

Keywords: Bayesian Inference, Bayesian Nonparametric Models, Gaussian Processes, Joint Quantile Model, Linear Quantile Regression, Monotone Curves.

1 Introduction

Ever since the seminal work by Koenker and Bassett (1978), linear quantile regression models have provided a useful and popular alternative to the traditional linear regression models. The latter, which link the conditional mean of a response to a linear combination of covariates, fail to provide an adequate modeling platform when different parts of the conditional response distribution are suspected to change at different rates. This difficulty is particularly acute when scientific interest focuses on how the covariates affect the tails and other non-central parts of the conditional distribution. Such situations routinely arise in economics, health and environment studies where the tails of the response distribution constitute events of special interest.

Let $Q_Y(\tau \mid x) := \inf\{q : P(Y \leq q \mid X = x) \geq \tau\}$ denote the τ-th conditional quantile ($0 \leq \tau \leq 1$) of a response Y given a vector of covariates $X = x$. A linear quantile regression model for $Q_Y(\tau \mid x)$, at a given τ, specifies

$$Q_Y(\tau \mid x) = \beta_0(\tau) + x'\beta(\tau) \tag{1}$$

where $\beta_0(\tau)$ is a scalar intercept, $\beta(\tau)$ is a coefficient vector of length $p = \dim(x)$ and x' denotes vector transpose of x. This specification retains the interpretability of linear

*Department of Statistical Science, Duke University, Durham, NC, mailto:st118@stat.duke.edu
†Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, mailto:kadane@stat.cmu.edu

regression by entertaining unknown parameters as linear coefficients. But, crucially, it targets a specific part of the conditional distribution of Y, encoded by the quantile point τ chosen by the analyst. By choosing τ appropriately, one can focus on the tails of the conditional distribution, as well as its other central or non-central parts. More importantly, by considering (1) simultaneously for all $\tau \in [0, 1]$, one obtains a complete description of the conditional distribution of Y (subject to monotonicity constraints that we discuss later) with the flexibility that x can have different effects on different parts of this distribution. The traditional linear regression model is a special case corresponding to a constant $\beta(\tau) \equiv \beta$.

For a given τ, Koenker and Bassett (1978) proposed to estimate the coefficients in (1) from data $\{(x_i, y_i) : 1 \le i \le n\}$ by minimizing the loss function $\sum_{i=1}^{n} \epsilon_i(\tau - I(\epsilon_i < 0))$ where $\epsilon_i = y_i - \beta_0(\tau) - x_i'\beta(\tau)$. This approach, which can be efficiently computed by linear programming, remains popular. A huge literature has emerged studying frequentist asymptotic properties of the resulting estimate, estimation of its standard error, derivation of tests of a given asymptotic size as well as various other extensions and improvements (see Koenker and Bassett 1978; Gutenbrunner and Jurečkóva 1992; Gutenbrunner et al. 1993; Koenker and Xiao 2002; Zhou and Portnoy 1996; Koenker and Machado 1999; Koenker 2005, and the references therein). This approach also influenced early attempts at a Bayesian analysis of (1) with a conditional sampling density for the response constructed as $Y = \beta_0(\tau) + x'\beta(\tau) + \sigma\varepsilon$ with ε having the asymmetric Laplace density $f_\varepsilon(\epsilon) = \text{const} \times \exp[-\epsilon(\tau - I(\epsilon < 0))]$ (Yu and Moyeed 2001; Tsionas 2003). Subsequent Bayesian approaches have looked into more flexible formulations of $f_\varepsilon(\epsilon)$, including non-parametric formulations with extensions to the heteroscedastic case: $f_\varepsilon(\epsilon \mid X = x)$ (Kottas and Gelfand 2001; Gelfand and Kottas 2003; Kottas and Krnjajić 2009; Thompson et al. 2010).

Arguably, a single-τ fit of (1) does not do justice to the full potential of the model, which lies in the simultaneous description

$$\{Q_Y(\tau \mid x) = \beta_0(\tau) + x'\beta(\tau); \;\; 0 \le \tau \le 1\} \tag{2}$$

encoded by the function valued parameters $\beta_0(\cdot)$ and $\beta(\cdot)$. A post-estimation pooling of the individual estimates, though valid from the viewpoint of asymptotic, frequentist calculations, faces serious philosophical and practical difficulties in drawing inference on $\beta(\tau)$ (or $\beta_0(\tau)$) simultaneously for a range of τ values when limited data are available. This is illustrated in Figure 1 that describes individual fits of (1) to a dataset on north Atlantic hurricane intensities (Elsner et al. 2008) with Y = maximum windspeed of a hurricane and X = year of its occurrence (between 1981 and 2006). The vertical bars in Figure 1 give the P-value for testing $H_0 : \beta(\tau) = 0$, as derived from the corresponding individual fits for $\tau \in \{0.01, 0.02, \cdots, 0.99\}$. It is not clear how to combine these P-values to draw inference on $\beta(\tau)$ even for fairly short ranges of τ values, say $\tau \in (0.4, 0.6)$. Moreover, a substantial fluctuation between the P-values highlights a poor borrowing of information across cases and indicates possible gaps in utilizing the information in the data in deriving a joint inference on the $\beta(\cdot)$ curve.

The pooling of single-τ fits also poses a serious foundational challenge to a Bayesian analysis of (1). The sampling distribution of Y under a Bayesian quantile regression

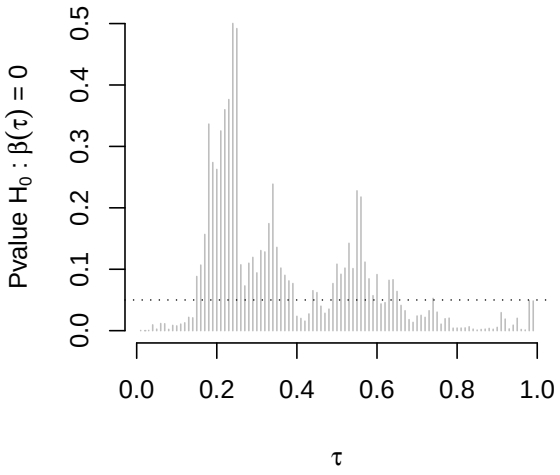

Figure 1: P-values from individual linear quantile regression analyses of north Atlantic tropical cyclone intensity against time. A substantial fluctuation leads to difficulties in drawing a composite inference.

model for a $\tau = \tau_1$ is usually different from that under the model for a different $\tau = \tau_2$. Therefore, although one could calculate posterior probabilities such as $P(\beta(\tau) < 0 \mid \text{data})$ for every individual τ-model, there is no coherent way to combine these probabilities across τ values, as they each represent posterior belief under a different model. Therefore, the issue of simultaneous inference remains unsolved (see, however, Dunson and Taylor 2005; Lancaster and Jun 2010, who offer partial solutions based on *pseudo* and *empirical* likelihoods).

A major obstacle in performing a simultaneous fitting of the joint model (2) appears to be the monotonicity constraint that the map $\tau \mapsto Q_Y(\tau \mid x)$ must be increasing in τ (non-decreasing if the distribution of Y has atoms) for every $x \in \mathcal{X}$, the domain of X. This constraint puts stringent restrictions on the $(\beta_0(\cdot), \beta(\cdot))$ curves that do not sit well with the loss function minimization approach of Koenker and Bassett (1978). It is possible to avoid the monotonicity problem altogether by specifying a nonparametric model for the conditional distribution $F_Y(y \mid x)$ and then inverting it to derive conditional quantile curves $Q_Y(\tau \mid x)$ (Scaccia and Green 2003; Geweke and Keane 2007; Taddy and Kottas 2010). The resultant curves, however, lack the interpretability of the linear model (2). This lack of interpretability could be a serious issue in studies where linear coefficients have meanings as rates of change with respect to input variables, such as time or diet, that can be understood and interpreted by an expert.

In this paper we introduce a semi-parametric Bayesian framework for a simultaneous analysis of (2). We make use of the observation that (2) automatically lends itself to a likelihood based inference on the function valued parameters $\beta_0(\cdot)$ and $\beta(\cdot)$. This is because, when $(\beta_0(\cdot), \beta(\cdot))$ are such that $\tau \mapsto Q_Y(\tau \mid x)$ is strictly increasing for each $x \in \mathcal{X}$, (2) uniquely determines a conditional sampling density for Y in the form:

$$f_Y(y \mid x) = \frac{1}{\frac{\partial}{\partial \tau} Q_Y(\tau \mid x)}\Big|_{\tau=\tau_x(y)} = \frac{1}{\frac{\partial}{\partial \tau}\beta_0(\tau) + x' \frac{\partial}{\partial \tau}\beta(\tau)}\Big|_{\tau=\tau_x(y)} \tag{3}$$

where $\tau_x(y)$ solves $y = Q_Y(\tau \mid x)$ in τ. A likelihood function over the monotonicity preserving choices of $(\beta_0(\cdot), \beta(\cdot))$ is then simply defined as $\prod_{i=1}^n f_Y(y_i \mid x_i)$. To establish (3) observe that $f_Y(y \mid x)$ is the derivative at y of the conditional distribution function $y \mapsto F_Y(y \mid x)$, which is the inverse of $\tau \mapsto Q_Y(\tau|x)$ and note that for any invertible and differentiable function $h(u)$ with derivative $\dot{h}(u)$, one has $\frac{\partial}{\partial v} h^{-1}(v) = 1/\{\dot{h}(h^{-1}(v))\}$.

For a univariate X, we show that the monotonicity constraint is easily satisfied by reparametrizing $\beta_0(\cdot)$ and $\beta(\cdot)$ as linear combinations of two monotonically increasing curves. Thus a Bayesian model is obtained whenever a prior distribution is specified on these monotone curves. We introduce a specific choice of this prior distribution induced by a logistic transformation of a smooth Gaussian process. An efficient Markov chain Monte Carlo (MCMC) is available for this model, making use of the recent approximation techniques developed in Tokdar (2007), Banerjee et al. (2008) and Tokdar et al. (2010).

For a multivariate X, a single index extension of the univariate model is proposed, i.e., the univariate model is applied to the one dimensional summary $X'\alpha$ where α is taken to be an additional model parameter. Although the single index approach captures only a subset of the monotonicity preserving joint models (2), it strikes a useful balance between computational difficulty and model richness. In particular, our single index implementation is computationally as efficient as a Bayesian implementation of the first order heteroscedastic model of He (1997) (as done in Reich et al. 2010), but provides non-trivial and measurable improvement over this simplistic approach that restricts the conditional densities $f_Y(y \mid x)$ to be linear location-scale transformations of each other. On the other hand, it is computationally easier than the recent proposal in Reich et al. (2011) which offers greater model coverage. Nevertheless, we put forward our multivariate model as only a promising extension for further research.

We present two illustrative real data applications of our Bayesian model. Our first application is an analysis of the north Atlantic hurricane data (Elsner et al. 2008) that was referred to in Figure 1. For this study, scientific interest focuses on how hurricane intensities are changing with time. The linear quantile regression model provides a suitable framework to study these changes in the form of linear trends, where the rate of change can be different in different parts of the intensity distribution. The upper tail is of particular interest due to the damage potential of the strongest hurricanes as well as their alleged connection with global warming (Trenberth 2005). We find that the trends estimated by our Bayesian model provide a natural smoothing of those found from the Koenker-Bassett scheme. The smoothing is a byproduct of borrowing of information across τ, which leads to a more homogeneous inference on how different parts

of the hurricane intensity distribution are changing with time. Our second application is a study of the relationship between infant birthweight and various demographic and pregnancy related factors of the mother (Abrevaya 2001). We illustrate that our single index model offers a smooth, homogeneous inference on this relationship across τ, but captures interesting features at the two tails by going beyond first order heteroscedasticity. Compared to individual Koenker-Bassett fits, our method provides equally accurate out-of-sample prediction without quantile crossing.

2 Linear Quantiles: The Univariate Case

2.1 The Model

For a univariate X, we can assume without loss of generality that $\mathcal{X}$ is bounded and convex, i.e., $\mathcal{X}$ is a bounded interval on the line. We shall take this interval to be $[-1, 1]$ with suitable redefinition of origin and scale, if necessary. Assuming $\mathcal{X}$ to be bounded is unavoidable for a valid linear specification of $Q_Y(\tau \mid x) = \beta_0(\tau) + x\beta_1(\tau)$, because the only non-intersecting lines on an unbounded $\mathcal{X}$ are parallel lines. Convexity of $\mathcal{X}$ is not restrictive, because lines that do not intersect each other over a given set remain non-intersecting also over its convex hull. An easy characterization of the required monotonicity of the quantile regression lines is offered by the following result.

Theorem 1. *A linear specification $Q_Y(\tau \mid x) = \beta_0(\tau) + x\beta(\tau)$, $\tau \in [0, 1]$ is monotonically increasing in τ for every $x \in \mathcal{X} = [-1, 1]$ if and only if*

$$Q_Y(\tau \mid x) = \mu + \gamma x + \frac{1-x}{2}\eta_1(\tau) + \frac{1+x}{2}\eta_2(\tau) \tag{4}$$

where $\eta_1(\tau)$ and $\eta_2(\tau)$ are monotonically increasing in $\tau \in [0, 1]$.

Proof. If $Q_Y(\tau \mid x)$ is given by (4) then it must be monotonically increasing in τ for every $x \in [-1, 1]$ for which both $(1 - x)/2$ and $(1 + x)/2$ are non-negative. One can express such a $Q_Y(\tau \mid x)$ as in (2) by defining $\beta_0(\tau) = \mu + (\eta_1(\tau) + \eta_2(\tau))/2$ and $\beta(\tau) = \gamma + (\eta_2(\tau) - \eta_1(\tau))/2$. For the converse, any monotonicity obeying $Q_Y(\tau \mid x)$ of the form (2) can be expressed as (4) by taking $\mu = 0$, $\gamma = 0$, $\eta_1 = Q_Y(\cdot \mid -1)$ and $\eta_2 = Q_Y(\cdot \mid 1)$. □

Therefore a model over $\eta = (\eta_1, \eta_2)$ induces via (4) a model over all valid, linear specifications of $Q_Y(\tau \mid x)$, provided it satisfies the boundary conditions:

$$Q_Y(0 \mid x) = \underline{y}(x), Q_Y(1 \mid x) = \overline{y}(x), \forall x \in \mathcal{X} \tag{5}$$

where $(\underline{y}(x), \overline{y}(x))$ gives the range of Y given $X = x$. We restrict attention to the special case where this range does not change with x, $\underline{y}(x) \equiv \underline{y}$, $\overline{y}(x) \equiv \overline{y}$, although linear changes are not difficult to accommodate. We allow both finite and infinite values for these boundaries.

To construct η_1, η_2 that are monotonically increasing and satisfy the above boundary conditions, we make use of monotonically increasing maps ξ_1, ξ_2 from $[0,1]$ onto itself and subject these to the following parametric transformations:

$$\eta_1 = \sigma_1 \tilde{Q}(\xi_1(\tau)), \quad \eta_2 = \sigma_2 \tilde{Q}(\xi_2(\tau)) \tag{6}$$

where $\sigma_1 > 0, \sigma_2 > 0$ and $\tilde{Q}(\tau)$ gives the quantiles of some fixed density with $\tilde{Q}(0) = \underline{y}$ and $\tilde{Q}(1) = \overline{y}$. For $\underline{y} = -\infty$, $\overline{y} = \infty$, one can take $\tilde{Q}$ to give the conditional quantiles of a $\mathsf{N}(\mu_0, \sigma_0)$ density or more generally, of a $\mathsf{t}_\nu(\mu_0, \sigma_0)$ density if heavy tails are desired. In this case $\mu, \gamma, \sigma_1, \sigma_2$ can be arbitrary and are treated as model parameters. When both $\underline{y}$ and $\overline{y}$ are finite, we take $\tilde{Q}$ to give the quantiles of a distribution supported over $[\underline{y}, \overline{y}]$ and fix $\mu = \gamma = 0$ and $\sigma_1 = \sigma_2 = 1$. In (6), $\tilde{Q}$ represents the target parametric model. Indeed, the parametric first order heteroscedastic model (He 1997; Reich et al. 2010) determined by linear location-scale changes of $\tilde{Q}$ is a special case when ξ_i's are the identity maps $\xi_i(\tau) = \tau$, $i = 1, 2$. Below we discuss a specific construction of $\xi = (\xi_1, \xi_2)$ where the identity map represents a central value for the ξ_i's.

Let $\omega(i, \tau)$ denote a zero-mean Gaussian process defined on $\{1, 2\} \times [0, 1]$, with covariance $\mathsf{Cov}(\omega(i, \tau), \omega(i', \tau')) = \kappa^2 c_{ii'} \exp(-\lambda^2(\tau - \tau')^2)$, where $c_{11} = c_{22} = 1$ and $c_{12} = c_{21} = \rho \in [0, 1]$, $\lambda > 0$, $\kappa > 0$ are to be specified later. Define

$$\xi_i(\tau) = \frac{\int_0^\tau e^{\omega(i,t)} dt}{\int_0^1 e^{\omega(i,t)} dt}, \ \tau \in [0, 1], i = 1, 2. \tag{7}$$

Then ξ_1, ξ_2 are monotonically increasing random functions that map $[0, 1]$ to $[0, 1]$. The transformation in (7), often called the logistic transformation, has been studied previously for modeling random densities (Lenk 1988, 1991, 2003; Tokdar 2007). Of importance to us is the result in Tokdar and Ghosh (2007) that the support of $\omega(i, \tau)$ includes all continuous or piecewise continuous functions. Due to continuity of the logistic transformation, the same can be said about ξ_i in supporting all continuous or piecewise continuous, monotonically increasing bijections of $[0, 1]$ onto itself. The zero-mean property of ω implies that ξ_i concentrates around the identity function – the logistic transform of the zero function.

To summarize, our specifications (4), (6) and (7) together define a valid, linear model on $Q_Y(\tau \mid x)$, with $\tilde{Q}$ as the base quantile function, supporting any continuous or piecewise continuous specification. It is easy to see that the first order heteroscedastic model is a special case of our specification with $\rho = 1$, with further reduction to the homoscedastic model with $\sigma_1 = \sigma_2$. We complete the model by specifying

$$(\rho, \lambda^2, \kappa^{-2}) \sim \mathsf{U}(0, 1) \times \mathsf{Ga}(5, 1/10) \times \mathsf{Ga}(3, 1/3), \tag{8}$$

although the analyses reported in the subsequent sections are fairly robust to the choice of prior on these parameters. A uniform prior on ρ offers a broad range of dependence between the curves ξ_1 and ξ_2 and puts positive mass around the first order heteroscedastic case $\rho = 1$. Our prior on λ specifies a central 95% interval $(0.36, 0.85)$ for the correlation between $\omega(i, \tau)$ and $\omega(i, \tau + 0.1)$ – a sufficiently wide range that precludes functions that are either too spiky or too flat but non-zero. A shape of 3 for the gamma distribution on κ^{-2} ensures a finite variance for the marginal distribution of $\omega(i, \tau)$.

2.2 Model Fitting

We handle the function valued variables $\omega(1,\tau)$ and $\omega(2,\tau)$ by approximating their domain $[0,1]$ with a dense grid $\{t_l = l\delta : l = 0,1,\cdots,L\}$. We used $L = 100$, $\delta = 0.01$ for the implementations reported below. During every iteration of the MCMC, $\zeta_i(\tau) = e^{\omega(i,\tau)}$, $i = 1,2$, are computed and stored only at the grid points $\tau = t_l$, $l = 0,1,\cdots,L$. These stored values are then used to compute numerically the integrals $\int_0^\tau \zeta_i(u)du$ for every τ on the grid, by using the trapezoidal rule. These integrals are stored and are used to get a trapezoidal approximation $\hat{\xi}_i(\tau)$ of $\xi_i(\tau)$ for every τ on the grid. The trapezoidal rule can be interpreted as performing the exact integration with an approximate $\hat{\zeta}_i(\tau)$ which equals $\zeta_i(\tau)$ at the grid points $\tau = t_l$ and equals the linear interpolation $\hat{\zeta}_i(\tau) = \{(\tau - t_{l-1})\zeta_i(t_l) + (t_l - \tau)\zeta_i(t_{l-1})\}/(t_l - t_{l-1})$ for a $\tau \in (t_{l-1}, t_l)$. This interpretation leads to an analytical evaluation of $\hat{\xi}_i(\tau) = \{(\tau - t_{l-1})\xi_i(t_l) + (t_l - \tau)\xi_i(t_{l-1}) - (\tau - t_{l-1})(t_l - \tau)(\zeta_i(t_l) - \zeta_i(t_{l-1})\}/(t_l - t_{l-1})$, for any $\tau \in (t_{l-1}, t_l)$, whenever such an evaluation is needed. It is this trapezoidal approximation $\hat{\xi}_i(\tau)$ that we use in (6) for the purpose of model fitting.

At the crux of our model fitting is the computation of the log-likelihood function

$$\begin{aligned}\sum_i \log f_Y(y_i \mid x_i) &= -\sum_i \log \frac{\partial}{\partial\tau} Q_Y(\tau_{x_i}(y_i)|x_i) \\ &= -\sum_i \log\left(\frac{1-x_i}{2}\frac{\partial}{\partial\tau}\eta_1(\tau_{x_i}(y_i)) + \frac{1+x_i}{2}\frac{\partial}{\partial\tau}\eta_2(\tau_{x_i}(y_i))\right) \quad (9)\end{aligned}$$

where $\tau_{x_i}(y_i)$ solves $y_i = Q_Y(\tau \mid x_i)$ in τ, $i = 1,2,\cdots,n$. A solution $\tau_x(y)$ to $Q_Y(\tau \mid x) - y = 0$ can be efficiently obtained by using Newton's recursion:

$$\tau_x^{(k+1)}(y) = \tau_x^{(k)}(y) - \frac{Q_Y(\tau_x^{(k)}(y) \mid x) - y}{\frac{\partial}{\partial\tau} Q_Y(\tau_x^{(k)}(y) \mid x)},$$

where $\tau_x^{(0)}(y)$ is some initial guess in $[0,1]$. Running this recursion would require repeated evaluations of $Q_Y(\tau \mid x)$ and $\frac{\partial}{\partial\tau}Q_Y(\tau \mid x)$ at various values of $\tau \in [0,1]$, which can be done relatively easily by using the trapezoidal approximations $\hat{\xi}_1(\tau), \hat{\xi}_2(\tau)$. Alternatively, one can simply search through the grid points to identify the interval (t_{l-1}, t_l) that contains $\tau_x(y)$ and approximate $\frac{\partial}{\partial\tau}Q_Y(\tau_x(y) \mid x)$ by $\{Q_Y(t_l \mid x) - Q_Y(t_{l-1} \mid x)\}/(t_l - t_{l-1})$.

The steps described in the above two paragraphs offer a fast algorithm to compute the log-likelihood at any given value of the parameter $\omega(\cdot,\cdot)$. This algorithm scales linearly in the number of observations n as well as in the number of grid points L. We were able to perform approximately 800 likelihood evaluations per second for the north Atlantic hurricane data presented in the next section, with $n = 291$ and $L = 100$, on a laptop computer with a 3.06 GHz Intel Core 2 Duo processor and 8 GB memory.

With an efficient algorithm in place to evaluate the log-likelihood function, we use Markov chain Monte Carlo to sample from and summarize the posterior distribution

of $(\omega, \lambda, \rho, \mu, \gamma, \sigma_1^2, \sigma_2^2)$ given data (κ^2 can be integrated out). Markov chain updating of (ω, λ, ρ) can be sticky due to the high degree of dependence between these variables as well as the need to invert large covariance matrices. A sparse surrogate that has been successfully implemented in the context of density estimation (Tokdar 2007) and spatial statistics (Banerjee et al. 2008), replaces ω with a knot-based approximation $\omega^*(i,\tau) = \mathsf{E}[\omega(i,\tau) \mid W^* := \{\omega(j,\tau_k^*) : j = 1,2, k = 0,1,\cdots,K\}]$, where K is a pre-specified order of approximation and $\{\tau_k^* = k/K : 0 \le k \le K\}$ is a set of knots over $[0,1]$. In our applications we use $K = 10$. The surrogate process ω^* can be easily evaluated at any τ given (W^*, λ, ρ) and these parameters along with the remaining parameters $(\mu, \gamma, \sigma_1^2, \sigma_2^2)$ are easily updated via a block-Metropolis sampler (R codes available at the first author's website: `http://www.stat.duke.edu/st118/~Software/`).

3 Application to Cyclone Intensity

Elsner et al. (2008) argue that the strongest tropical cyclones in the North Atlantic basin have gotten stronger over the last couple of decades. Their analysis includes fitting separate linear quantile regression models to maximum sustained wind speed (`WmaxST`) of tropical cyclones (including tropical storms) against their year of occurrence (`Year`) over a range of τ values in $[0,1]$. The slopes of these regression lines are found statistically different from zero (with positive estimated values) for some of the chosen τ values, mostly in the upper tail $\tau > 0.75$, leading to the use of the qualifier *strongest* in their summary. Figure 1 shows the P-values corresponding to these tests[1] for τ on the grid $\{0.01, 0.02, \cdots, 0.99\}$.

In this section we present an analysis of the data used by Elsner et al. (2008) with the joint quantile regression model discussed in the previous section. We consider $\underline{y} = 0$, $\overline{y} = \infty$ for the range of `WmaxST`, measured in nautical miles per hour (knots), and restrict $\tilde{Q}$ to match the quantiles of a power-Pareto density

$$\tilde{f}(y) = \frac{ak}{\sigma} \frac{k(\frac{y}{\sigma})^{k-1}}{(1+(\frac{y}{\sigma})^k)^{-(a+1)}}, \quad y > 0. \tag{10}$$

For a random variable Y with (10) as its density function, $(Y/\sigma)^k$ has the familiar Pareto distribution with index a. The heavy right tail of the Pareto distribution ensures that $\tilde{f}$ entertains occasional occurrences of extremely strong tropical cyclones while the power transformation with a $k > 1$ makes $\tilde{f}(y)$ vanish quickly as $y \to 0$. One could argue that $\underline{y}$ should be fixed at 35, as a storm system is required to have maximum sustained winds of at least 35 knots to be labeled as a tropical cyclone. However, this thresholding applies to the best-track record of maximum wind, while Elsner et al. (2008) derive `WmaxST` only from satellite data. The two measurements are close but not identical, and some of the storms in the data set had maximum wind below 35 knots.

We fix $a = 0.45, \sigma = 52$ and $k = 4.9$. The corresponding $\tilde{f}$ well approximates the median and the interquartile range of the best-track maximum winds of all tropical

[1] Obtained through the `rq()` function of the `R` package `quantreg`, a 10000 Bootstrap sample is used to compute P-values.

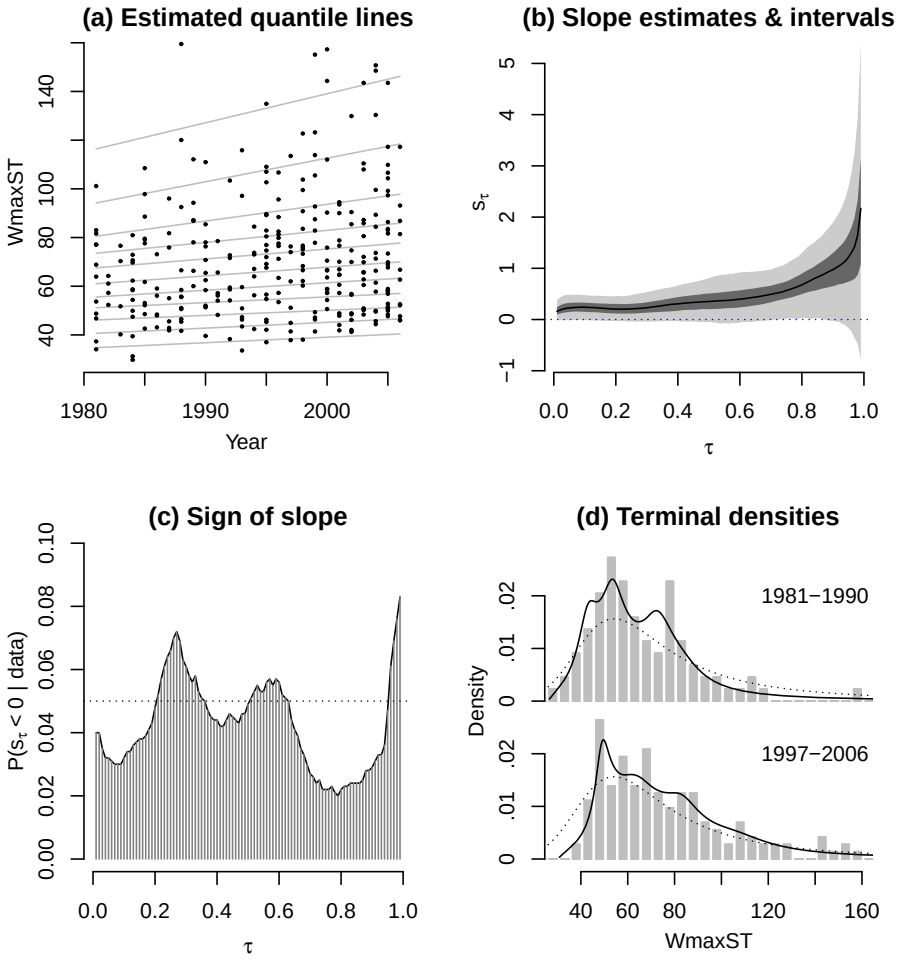

Figure 2: Posterior summaries of our joint quantile regression analysis of maximum wind speed (`WmaxST`) of north Atlantic tropical cyclones against year (`Year`) of occurrence. (a) Posterior mean of $Q_Y(\tau \mid x)$ for $\tau \in \{0.05, 0.1, 0.2, \cdots, 0.9, 0.95\}$ overlaid on data scatter. (b) Posterior medians and 50% and 95% central credible intervals for slopes $s_\tau = \frac{\partial}{\partial x} Q_Y(\tau \mid x)$. (c) Posterior probability of s_τ being negative. (d) Terminal conditional densities $f_Y(y \mid 1981)$ and $f_Y(y \mid 2006)$ (solid lines) found by inverting posterior means of $Q_Y(\tau \mid 1981)$ and $Q_Y(\tau \mid 2006)$, overlaid on the histograms of `WmaxST` pooled over the first and the last 10 years of study. The dashed curves are the base power-Pareto density $\tilde{f}$.

cyclones[2] between 1900-1979. We proceed with the model in (6) with $\mu = 0$ and $\sigma_1^2, \sigma_2^2 \overset{\text{IID}}{\sim} \mathsf{Ga}(2,2)$ which ensures each σ_i^2 is centered around 1, but with a wide spread. The ratio of σ_1 to σ_2 has prior median 1, and has 80% chance to be between 1/3 and 3.

The data scatter is shown in Figure 2(a), where each point represents a tropical cyclone between 1981 and 2006, with `Year` on the horizontal axis and `WmaxST` on the vertical axis. Overlaid on the scatter plot are some of the quantile lines estimated from our joint model. Figure 2(b) shows the posterior credible intervals of the slopes $s_\tau = \frac{\partial}{\partial x} Q_Y(\tau \mid x)$ and Figure 2(c) shows the posterior probabilities $P(s_\tau < 0 \mid \text{data})$. The solid lines in Figure 2(d) are the two terminal conditional densities $f_Y(y \mid 1981)$ and $f_Y(y \mid 2006)$ obtained by inverting, via (3), the posterior mean estimates of $Q_Y(\tau \mid 1981)$ and $Q_Y(\tau \mid 2006)$. For a visual comparison, we have included in these plots histograms of `WmaxST` pooled over the first and the last ten years of the study. The dashed lines give the density $\tilde{f}(y)$. All posterior summaries appearing in Figure 2 and below are Monte Carlo approximations based on a sample of 1000 parameter values that we obtained by running a block-Metropolis sampler for 100,000 iterations, discarding the first 10,000 iterations and saving every 90-th draw from the remaining iterations. Trace plots of s_τ (not shown here) exhibit no drifts and the autocorrelation of successive saved draws s_τ drops below 0.1 at lag 1, for all τ.

Figures 2(a)-(c) clearly indicate an upward trend of `WmaxST` across the entire range of $\tau \in [0,1]$. Indeed, the posterior probabilities of s_τ being positive, which equal $1 - P(s_\tau < 0 | \text{data})$, are estimated to be 92% or more for all τ between $[0.01, 0.99]$. This indicates that TCs in almost the entire range of intensity distribution have gotten stronger during 1981–2006. This is in contrast to the report of Elsner et al. (2008) who fail to conclude increasing trends except for TCs in the upper tail of the intensity distribution (see also Figure 1). It is, however, interesting that the estimated values of s_τ, $\tau \in (0,1)$ from our analysis are very similar to those found by the classical, separate τ quantile regression analysis as used in Elsner et al. (2008).

The joint quantile regression analysis presented above is quite robust to the choice of parameters, including that of the base density $\tilde{f}$. However, inference in the tails is mildly sensitive to the tails of $\tilde{f}$. This is not surprising since data are sparse in the tails, leaving the prior with a bigger influence on the posterior. For illustration, we considered an 'all purpose' choice of $\tilde{f}$ given by the $\mathsf{t}_1(100, 8)$. This choice is not entirely suitable for `WmaxST` as it gives $\underline{y} = -\infty$. However, our choice of center and scale ensures that the central 95% interval of the base density $\tilde{f}$ is given by (0, 200). We again consider the model in (6), with $\mu \sim \mathsf{t}_1(0,1)$, $\gamma \sim \mathsf{t}_1(0,1)$ and $\sigma_1^2, \sigma_2^2 \overset{\text{IID}}{\sim} \mathsf{Ga}(2,2)$. Upon fitting this model to data, we find that the estimated quantile lines and the credible intervals on slopes (Figure 3(a)) are quite similar to those found in our previous analysis with the power-Pareto base model, except in the extreme tails where quantiles are estimated to be steeper than before. The estimated posterior probabilities of s_τ being negative also remain mostly unchanged, except in the extreme lower tail. It should be noted that the current base model differs from the power-Pareto model most severely in the lower tail.

[2]Source: http://weather.unisys.com/hurricane/atlantic/

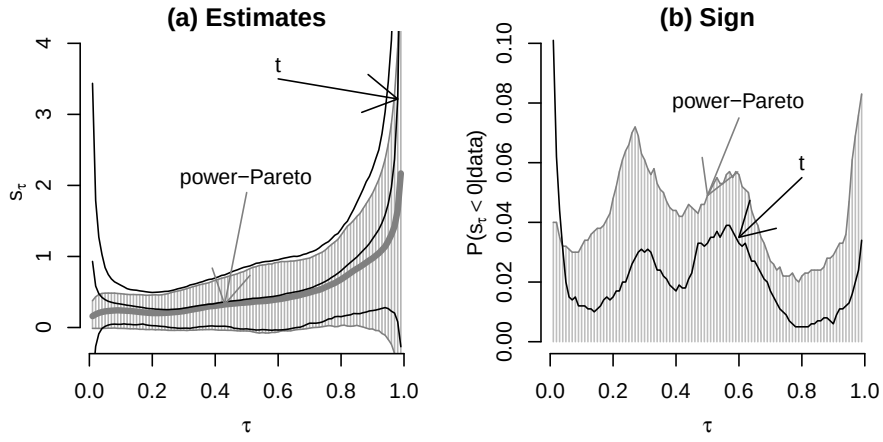

Figure 3: Posterior summaries of s_τ under two choices of the base density $\tilde{f}$: power-Pareto (gray lines) and $t_1(100, 8)$ (black lines). (a) Posterior medians and 95% credible intervals of s_τ. (b) Posterior probabilities of s_τ being negative.

Choosing the base to be a t-density instead of a power-Pareto density has another subtle effect on the inference on the quantile lines. The $\mathtt{t}(100, 8)$ base model produces a posterior that concentrates over quantiles curves that are heteroscedastic beyond the first order. A necessary condition for quantile curves to be first order heteroscedastic is that Δs_τ does not change its sign in the interior of $[0, 1]$. Therefore a lower bound on $P(Q_Y$ is heteroscedastic beyond first order $\mid$ data) is given by the posterior probability of a sign change of Δs_τ. We estimate this latter posterior probability to be 0.004 for the power-Pareto model and 0.798 for the t model. These are conservative estimates as we round Δs_τ to the nearest tenth before checking for a change in sign. It is apparent from Figure 3(a) that unlike the power-Pareto base, the t base model favors a change of sign in Δs_τ in the lower tail; s_τ first decreases in τ, then increases. This difference between the two models can be explained by their treatment of the lower tail. The t model, with an unbounded, heavy lower tail, supports steeper $Q_Y(\tau \mid x)$ for τ close to zero.

In Figure 4 we compare posterior summaries of s_τ from our joint, base-t model with those from a collection of single-τ fits of a Bayesian linear quantile model with asymmetric Laplace error distribution (Yu and Moyeed 2001). The base-t version is chosen because it accommodates an unbounded range for the response similar to the asymmetric Laplace model. At each $\tau \in \{0.01, \cdots, 0.99\}$, the corresponding asymmetric Laplace model is fitted with a flat prior $p(\beta_0(\tau), \beta(\tau), \sigma^2) \propto 1/\sigma^2$, however, the posterior is quite robust to the choice of prior. Posterior summaries remain virtually the same if instead normal or t distributions were used as prior distributions for $\beta_0(\tau)$ and $\beta(\tau)$ and a gamma or an inverse gamma prior with a small degrees of freedom was assigned

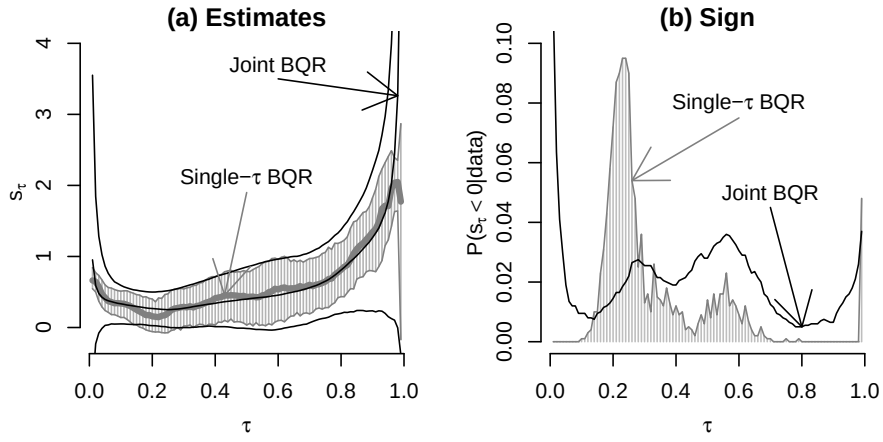

Figure 4: Comparison of joint and single-τ Bayesian quantile regression fits for tropical cyclone study. (a) Posterior medians and 95% credible intervals of s_τ. (b) Posterior probabilities that s_τ is negative. In both plots, gray lines show posterior summaries from the single-τ fits, and the black lines show posterior summaries from the joint fit.

to σ^2.

Figure 4(a) indicates that the estimates of s_τ, taken to be the posterior medians, are similar for the joint and single-τ fits. The curve of estimated s_τ, obtained by combining the single-τ fits, appears slightly more bumpy than the curve obtained from the joint fit. This is not surprising because the single-τ fits learn a piece of the curve at every τ, while the joint model deals with the whole curve simultaneously. The more striking difference lies in the credible intervals for s_τ, specially for τ values larger than 0.5 for which the credible intervals from the single-τ fits are always shorter than the same from the joint fit. It is also counterintuitive that the credible intervals from the single-τ fits get shorter as τ approaches 0 or 1. One would expect that due to sparse data at the tails the intervals ought to be larger. This is precisely the case with the joint fit.

This rather surprising behavior of the credible intervals from the single-τ fits can be explained as follows. The likelihood under an asymmetric Laplace model at a fixed τ is determined by how well the residuals $\epsilon_i = y_i - \beta_0(\tau) - \beta(\tau)x_i$ are explained by the error distribution f_ϵ. For τ close to 0 or 1, small changes in $\beta_0(\tau), \beta(\tau)$ can push many of these residuals far into the tails of the error distribution. Consequently, only a small range of these parameter values is entertained, producing shorter credible intervals for s_τ. The joint fit, on the other hand, has a much greater flexibility in accounting for uncertainty in s_τ in a localized way. The specification of $Q_Y(\tau \mid x)$ at a given τ value will have little dependence on the same at another τ value that is far apart. The lack of this ability of localized learning for a single-τ fit also manifests in the strong sensitivity of posterior inference on the choice of the error density f_ϵ. Posterior medians and credible intervals

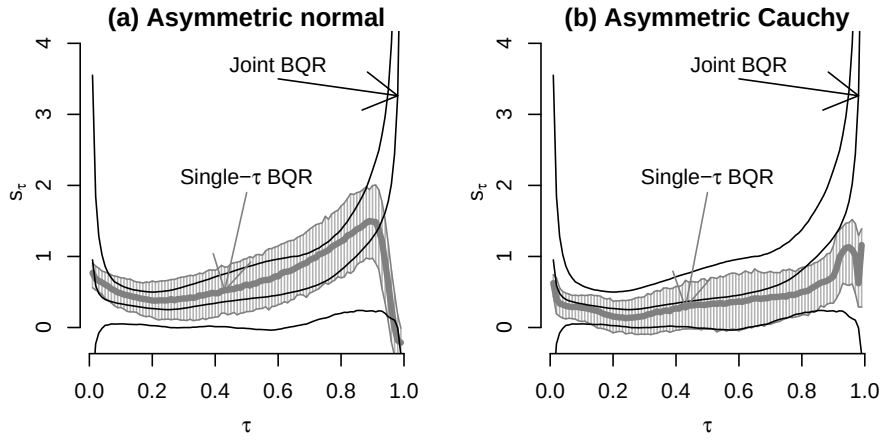

Figure 5: Sensitivity of the single-τ fits to the choice of error distribution. (a) Posterior medians and 95% credible intervals of s_τ under an asymmetric normal error distribution (gray lines). (b) The same under an asymmetric Cauchy error distribution (gray lines). The summaries from the base-t joint fit are used as a reference (black lines). For a $\tau \in [0,1]$, asymmetric Laplace, normal and Cauchy densities are defined by $f(\epsilon) = 2g_0(2\epsilon(\tau - I(\epsilon < 0)))$ with g_0 equalling, respectively, $g_0(x) = (1/2)\exp\{-|x|\}$, $g_0(x) = (1/\sqrt{2\pi})\exp\{-x^2/2\}$ and $g_0(x) = (1/\pi)\{1+x^2\}^{-1}$ over $x \in (-\infty, \infty)$.

of s_τ change sharply if f_ϵ is changed from an asymmetric Laplace density to a more thin tailed asymmetric normal density or to a more flat tailed asymmetric Cauchy density (Figure 5).

4 Linear Quantiles: The Multivariate Case

4.1 Model

By interpreting η_1 and η_2 in (4) as the conditional quantiles of $Y - \mu - \gamma X$ at $X = -1, 1$, one could build a similar construction for a multivariate X as follows. Fix $p+1$ linearly independent vectors $a_k = (1, \tilde{a}_k')'$ with $\tilde{a}_k \in \mathcal{X}$, $k = 1, \cdots, p+1$ and let $A = [a_1 : \cdots : a_{p+1}]$. Define

$$Q_Y(\tau \mid x) = \mu + x'\gamma + (1, x')(A')^{-1}\eta(\tau), \quad \tau \in [0,1], x \in \mathcal{X} \tag{11}$$

where $\eta(\tau) = (\eta_1(\tau), \cdots, \eta_{p+1}(\tau))$ with each $\eta_k(\tau)$ monotonically increasing in $\tau \in [0,1]$. It is easy to see that such $Q_Y(\tau \mid x)$ is monotonically increasing in $\tau \in [0,1]$ for every x in the convex hull of $\{a_1, \cdots, a_{p+1}\}$. This convex set, however, is often very small compared to $\mathcal{X}$ for moderately large p even for the best possible choice of $\{a_k\}$. Verifying monotonicity of $Q_Y(\tau \mid x)$ on the whole of $\mathcal{X}$, which is equivalent to verifying this

at the extreme points of the convex hull of $\mathcal{X}$, takes the form of an overdetermined system whenever the number of these extreme points exceeds $p+1$. For most choices of η, the monotonicity condition fails to hold, making the model specifications in (11) computationally intractable (see, however Reich et al. 2011, for an interesting alternative construction over a finite grid of τ values).

This motivates us to look for computationaly tractable alternatives to (11), possibly at the cost of the support of the model. One attractive choice is a single-index generalization of our univariate model, where the dimensionality of X is reduced to 1 by means of a linear projection of the covariate vector. For $\alpha \in \mathbb{R}^p$, $-\infty < a < \infty$ and $b > 0$, let $\mathcal{X}_{\alpha,a,b}$ denote the cylinder $\{x \in \mathbb{R}^p : x'\alpha \in (a-b, a+b)\}$ and $p_{\alpha,a,b}(x)$ denote the shifted and scaled projection $(x'\alpha - a)/b$. Define

$$Q_Y(\tau \mid x) = \mu + x'\gamma + \frac{1 - p_{\alpha,a,b}(x)}{2}\eta_1(\tau) + \frac{1 + p_{\alpha,a,b}(x)}{2}\eta_2(\tau), \quad x \in \mathcal{X}_{\alpha,a,b} \tag{12}$$

where $\eta_i(\tau) = \sigma_i \tilde{Q}(\xi_i(\tau))$, $i = 1, 2$ are exactly as in (6). Although (12) does not support all valid linear specifications of $Q_Y(\tau \mid x)$, it does embed as a special case the first order heteroscedastic model whenever $\xi_1 \equiv \xi_2$, which corresponds to $\rho = 1$. Furthermore, the projection vector α offers a global, single-index summary of the relative influence of the components of x.

We model α with a p-variate t-distribution: $\alpha \mid \sigma_\alpha^2 \sim \mathsf{N}(0, \sigma_\alpha^2 I_p)$, $1/\sigma_\alpha^2 \sim \mathsf{Ga}(1/2, 1/2)$, with the understanding that the component variables of x are of similar scales, which can be ensured by standardization of the observed values x_i. The cylinder edges a, b are fixed as: $a = (\max_i x'\alpha_i + \min_i x'\alpha_i)/2$ and $b = (\max_i x'\alpha_i - \min_i x'\alpha_i)/2$ to ensure every observed x_i is within the corresponding cylinder $\mathcal{X}_{\alpha,a,b}$. Such a data dependent choice is unavoidable as the knowledge of the convex hull where X lives is crucial in defining non-intersecting linear conditional quantiles.

4.2 Illustration with Birth Weight Data

As an illustration, we study the effect of a multitude of pregnancy related factors on infant birthweight (`BirthWt`, in grams) quantiles. A detailed analysis of this effect, as in Abrevaya (2001); Koenker and Hallock (2001), is beyond the scope of this paper. We rather focus on demonstrating various aspects of our model fit and compare the resulting inference with individual quantile regression fits and the corresponding frequentist inference. Our data consist of 5000 randomly selected entries from the June 1997 detailed natality records[3] of the United States on singleton, live births to mothers recorded as either black or white, in the age group 18-45. As covariates we include gender of the child (`Boy`, boy = 1, girl = 0), mother's age (`Age`, in years), average daily number of cigarattes during pregnancy (`Cigarette`) and weight gain during pregnancy (`WeightGain`, in pounds) and indicators for her being married (`Married`), black (`Black`), high school graduate (`HighSchool`), college dropout (`SomeCollege`), college graduate (`College`), without any prenatal care (`NoPrenatal`), with prenatal care from second

[3]Obtained from National Bureau of Economic Research: www.nber.org/natality/1997/

trimester onward (`PrenatalSecond`), from third trimester onward (`PrenatalThird`) and not a smoker (`NoSmoke`). `NoSmoke` is included to reflect the belief that a jump from zero cigarettes to one cigarette is fundamentally different from a unit increase when the mother is already a smoker.

We proceed with the model in (12), with $\tilde{Q}(\tau)$ giving the quantiles of the $\mathtt{t}_1(4000, 320)$ density which gives 95% mass to $(0, 8000)$. We take $\mu \sim \mathtt{t}_1(0, 320/3)$ and the same prior is used on each component of γ. The variance inflation parameters are modeled as $\sigma_1^2, \sigma_2^2 \overset{\text{IID}}{\sim} \mathtt{Ga}(2, 2)$. Although the model is fitted to the standardized versions of x_i's, the posterior summaries reported below correspond to the original origin and scale.

Figures 6(a)-(m) show posterior medians and central 95% credible intervals for the slope parameters $s_j(\tau) = \frac{\partial}{\partial x_j} Q_Y(\tau \mid x)$ from our model fit, overlaid on the estimated slopes and 95% confidence intervals obtained from individual fits. Except for variables relating to education level and to some extent prenatal care, every other variable appears to have an influence on the quantiles of birthweight. This influence is substantially non-constant across the quantiles for infant's gender and the mother's weight gain during pregnancy. While gender has a more pronounced effect in the middle range, weight gain contributes to substantially higher birthweight at both the low and the high ends of the weight distribution. Slopes of `Boy` and `WeightGain` clearly indicate heteroscedasticity beyond the first order, in fact, our Monte Carlo approximation to $P(Q_Y$ is heteroscedastic beyond the first order $\mid$ data) is exactly 1.

While the two sets of summaries are similar in appearance (ignoring smoothness), there is noticeable difference in the tails, particularly in the lower tail. The individual fits show more dramatic features in the lower tail than our joint fit. To assess whether this difference indicates a shortcoming of the single-index model in capturing the complex structure of the natality data, we compare its fit to that of the individual models on a new random sample of 5000 entries from the June 1997 records, with the same criteria applied on the mother as mentioned earlier. Figure 6(n) shows the graphs of average prediction errors $\frac{1}{5000}\sum_{i=1}^{5000} \epsilon_i^*(\tau)(\tau - I(\epsilon_i^*(\tau) < 0))$ against τ, for test data residuals $\epsilon_i^*(\tau) = y_i^* - \hat{Q}_Y(\tau \mid x_i^*)$ where $\hat{Q}_Y(\tau \mid x_i^*)$ equals the estimated value of $Q_Y(\tau \mid x_i^*)$ for the individual fits and equals the posterior mean of $Q_Y(\tau \mid x_i^*)$ for our single-index joint model. The two prediction error measures are virtually indistinguishable for $\tau \in [0.03, 0.97]$, with slightly inferior values for the single-index model at the extreme tails outside this range. These extreme tails are a little more elongated than what would have given an ideal fit, mostly because the heavy tails of the prior base $\tilde{Q}$ prevail over data in these regions. However, the joint fit comes with the advantage of interpretable quantile curves that do not intersect each other. Figures 6(o)-(p) show the quantile lines for a hypothetical mother whose weight gain is changed from 0 lb. to 100 lb., keeping all other attributes fixed at their corresponding average values as recorded in our data (`WeightGain` in the June 1997 natality records ranges from 0 to 98, with several cases in the nineties). The estimated lines from the individual fits intersect in the lower range of τ.

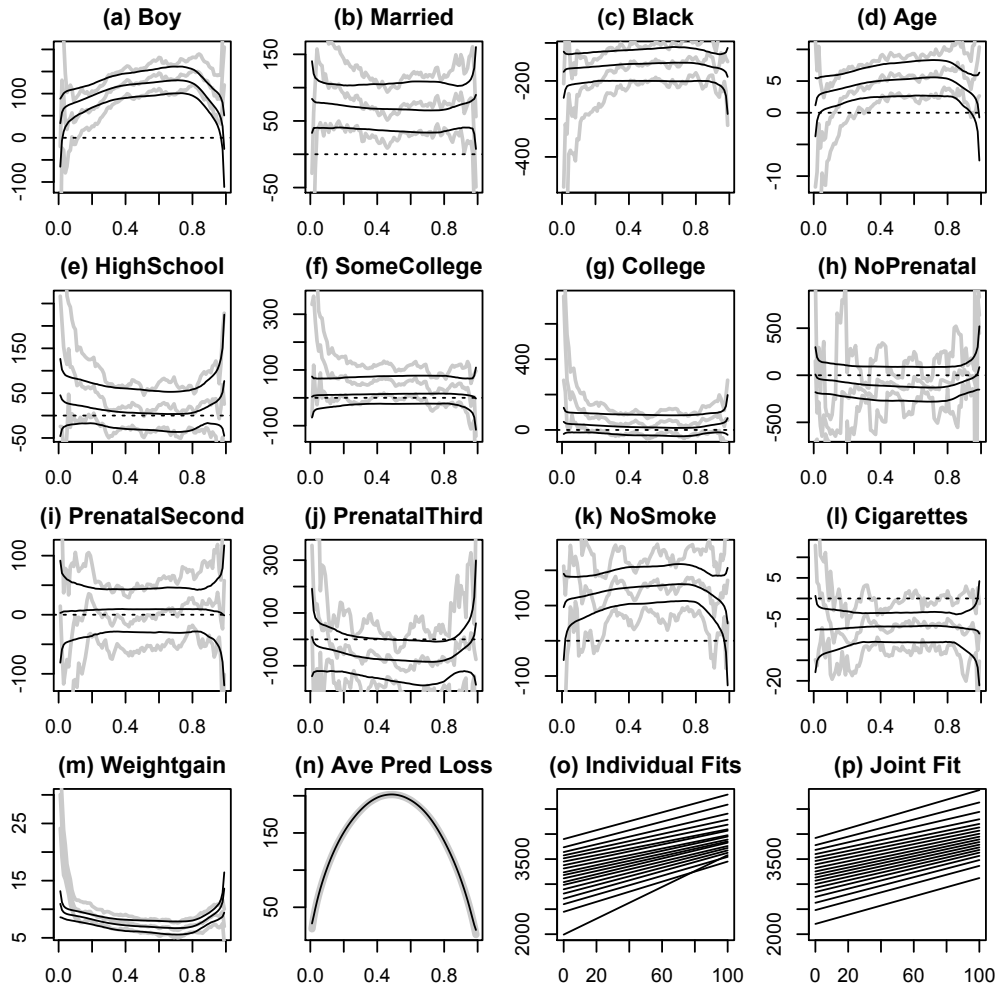

Figure 6: Quantile regression analysis of Birthweight data. (a)-(m) Posterior median and 95% credible intervals for slopes from the joint fit (thin black lines) overlaid on estimated values and 95% confidence intervals from individual fits (thick gray lines). (n) Average prediction error on test data for individual fits (thick gray line) and joint fit (thin black line). (o)-(p) Fitted quantile lines for a hypothetical mother whose weight gain is changed from 0 to 100 lbs with all other attributes kept at their average values recorded from the data. Fitted lines ($\tau \in \{0.05, 0.1, 0.15, \cdots, 0.95\}$) from individual fits cross in the lower tail.

5 Concluding Remarks

This paper presents a Bayesian framework for fitting the linear quantile regression model (2) simultaneously at all quantile points τ. The hurricane intensity analysis presented in Section 3 indicates the advantages of a simultaneous quantile regression analysis over individual fits. A simultaneous analysis has the flexibility to borrow information across cases and to better account for uncertainty in quantile line slopes in a localized way. The resulting difference in inference can have nontrivial implications on an overall summary and interpretations of results.

Multivariate quantile regression is considerably more difficult than the univariate case. Although our proposal in Section 4 provides a solution for the multivariate case, our work should be regarded as an initial exploration with admitted difficulties and caveats. Our analysis of the infant birthweight data (Section 4), however, shows that the single-index approach can match Koenker-Basett fits in predictive power, with tighter inference on model parameters. It is also evident that it generalizes the first order heteroscedastic model in a practically useful way. This additional flexibility does not require any extra validation of monotonicity and retains the interpretability of an ordinary linear model.

A crucial step in fitting our univariate model is a linear transformation of the original predictor variable into the range $[-1, 1]$. There is clearly some flexibility regarding this choice and the posterior inference is likely to depend on it. For our tropical cyclone analysis, we transformed `Year` so that the year 1981 was mapped to -1 and the year 2006 was mapped to 1. This amounts to assuming that the quantiles of the TC intensity distribution change linearly between 1981 and 2006, which is the same as the period under study. If instead the linearity assumption is made over a longer period, then we would map the observation period [1981, 2006] to a smaller subinterval of $[-1, 1]$. Figure 7 reports posterior summaries under two such cases. In one the linearity assumption is made over the period [1968, 2019], so that the observation period gets mapped to $[-0.5, 0.5]$. In the other the linearity assumption is made over the period [1929, 2058] and the observation period is mapped to $[-0.2, 0.2]$.

Note that when linearity is assumed over a longer period, the monotonicity constraint will restrict the model to entertain a smaller range of slope values. Figure 7 shows that this restriction has a mild effect on the posterior summaries of s_τ when the period of linearity is enlarged two fold from [1981, 2006] to [1968, 2019] and a substantial effect when enlarged five fold to [1929, 2058].

The logistic Gaussian process construction presented here is attractive both from computational and theoretical perspectives. Hjort and Walker (2009) have investigated Kullback-Leibler support conditions for Bayesian density models specified through quantiles. Their Proposition 3.1 holds verbatim for quantile functions defined via the logistic Gaussian process. However, generalizing this result to the quantile regression setting, possibly with an unbounded response variable, would require substantial further work.

The simultaneous linear quantile regression methodology presented here is just one of many ways of generalizing the simple linear regression. A very different way of achieving

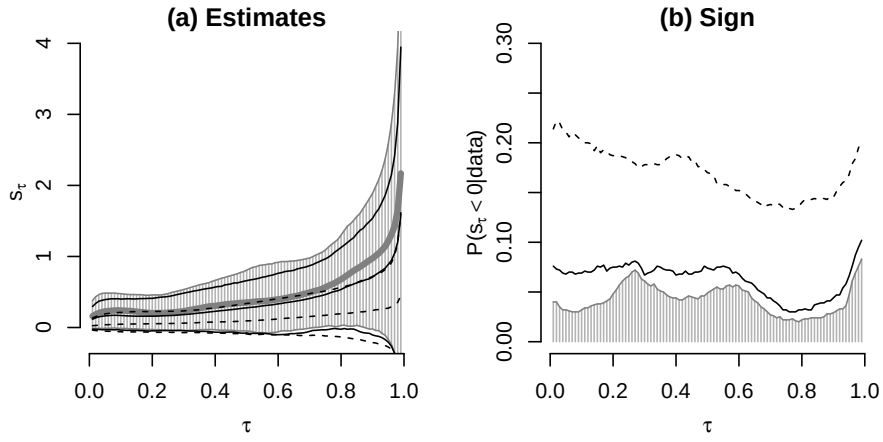

Figure 7: The effect of the length of the linearity period over posterior summaries of s_τ, for cyclone study. (a) posterior median and 95% credible intervals of s_τ. (b) Posterior probabilities that s_τ is negative. In each plot, gray lines show posterior summaries under the power-Pareto model of Section 3 with linearity assumed over [1981,2006], black solid lines show the same when linearity is assumed over [1968, 2019] and black dashed lines show the same when linearity is assumed over [1929,2058].

this generalization is to place a flexible model on the conditional density of the response variable. This approach has recently gained attention in the Bayesian literature with the introduction of a number of non-parametric and semi-parametric prior distributions on conditional densities (Müller et al. 1996; MacEachern 1999, 2000; Griffin and Steel 2006; Dunson et al. 2007; Tokdar et al. 2010). Whether to model conditional quantiles or to model conditional densities largely depends on one's subjective preferences. The two approaches need not always be different. A flexible nonparametric model on the conditional quantiles is equivalent to the same on the conditional density (Scaccia and Green 2003; Geweke and Keane 2007; Taddy and Kottas 2010). But, when the conditional quantiles are modeled under shape restrictions, such as the linearity conditions of our model, the two approaches become very different. The simultaneous linear quantile model is much less flexible in capturing changes in the conditional behavior of the response. But the changes it captures can be interpreted through well understood linear coefficients. The net worth of this trade-off between flexibility and interpretability can be debated.

Acknowledgement

We would like to thank the Editor, an Associate Editor and a referee whose feedback on an earlier draft of this paper has led to a much improved exposition of the material.

References

Abrevaya, J. (2001). "The effects of demographics and maternal behavior on the distribution of birth outcomes." *Empirical Economics*, 26: 247–257. 5, 14

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). "Gaussian predictive process models for large spatial data sets." *Journal of the Royal Statistical Society Series B*, 70(4): 825–848. 4, 8

Dunson, D. B., Pillai, N. S., and Park, J.-H. (2007). "Bayesian density regression." *Journal of the Royal Statistical Society B*, 69: 163–183. 18

Dunson, D. B. and Taylor, J. A. (2005). "Approximate Bayesian inference for quantiles." *Journal of Nonparametric Statistics*, 17: 385–400. 3

Elsner, J. B., Kossin, J. P., and Jagger, T. H. (2008). "The increasing intensity of the strongest tropical cyclones." *Nature*, 455: 92–95. 2, 4, 8, 10

Gelfand, A. E. and Kottas, A. (2003). "Bayesian Semiparametric Regression for Median Residual Life." *Scandinavian Journal of Statistics*, 30: 651–665. 2

Geweke, J. and Keane, M. (2007). "Smoothly mixing regression." *Journal of Econometrics*, 138(1): 252–290. 3, 18

Griffin, J. E. and Steel, M. F. J. (2006). "Order-Based Dependent Dirichlet Processes." *Journal of the American Statistical Association*, 101: 179–194. 18

Gutenbrunner, C. and Jurečkóva, J. (1992). "Regression rank-scores and regression quantiles." *The Annals of Statistics*, 20: 305–330. 2

Gutenbrunner, C., Jurečková, J., Koenker, R., and Portnoy, S. (1993). "Tests of linear hypotheses based on regression rank scores." *Journal of Nonparametric Statistics*, 2: 307–333. 2

He, X. (1997). "Quantile Curves without Crossing." *The American Statistician*, 51: 186–192. 4, 6

Hjort, N. D. and Walker, S. G. (2009). "Quantile pyramids for Bayesian nonparametrics." *The Annals of Statistics*, 37(1): 105–131. 17

Koenker, R. (2005). *Quantile Regression*. Cambridge U. Press. 2

Koenker, R. and Bassett, G. (1978). "Regression Quantiles." *Econometrica*, 46: 33–50. 1, 2, 3

Koenker, R. and Hallock, K. F. (2001). "Quantile Regression." *Journal of Economic Perspectives*, 15: 143–156. 14

Koenker, R. and Machado, J. (1999). "Goodness of Fit and Related Inference Processes for Quantile Regression." *Journal of the American Statistical Association*, 94(448): 1296 –1310. 2

Koenker, R. and Xiao, Z. (2002). "Inference on the Quantile Regression Process." *Econometrica*, 70(4): 1583–1612. 2

Kottas, A. and Gelfand, A. E. (2001). "Bayesian semiparametric median regression modeling." *Journal of the American Statistical Association*, 96: 1458–1468. 2

Kottas, A. and Krnjajić, M. (2009). "Bayesian nonparametric modeling in quantile regression." *Scandinavian Journal of Statistics*, 36: 297–319. 2

Lancaster, T. and Jun, S. J. (2010). "Bayesian quantile regression methods." *Journal of Applied Econometrics*, 25(2): 287–307. 3

Lenk, P. J. (1988). "The Logistic Normal Distribution for Bayesian, Nonparametric, Predictive Densities." *Journal of American Statistical Association*, 83: 509–516. 6

— (1991). "Towards a Practicable Bayesian Nonparametric Density Estimator." *Biometrika*, 78: 531–543. 6

— (2003). "Bayesian Semiparametric Density Estimation and Model Verification Using a Logistic Gaussian Process." *Journal of Computational and Graphical Statistics*, 12: 548–565. 6

MacEachern, S. M. (1999). "Dependent Nonparametric Processes." In *Proceedings of the Section on Bayesian Statistical Science*, 50–55. Alexandria, VA: American Statistical Association. 18

— (2000). "Dependent Dirichlet Processes." Ohio State University, Dept. of Statistics Technical Report. 18

Müller, P., Erkanli, A., and West, M. (1996). "Bayesian curve fitting using multivariate normal mixtures." *Biometrika*, 83: 67–79. 18

Reich, B. J., Bondell, H. D., and Wang, H. (2010). "Flexible Bayesian quantile regression for independent and clustered data." *Biostatistics*, 11(2): 337–352. 4, 6

Reich, B. J., Fuentes, M., and Dunson, D. (2011). "Bayesian spatial quantile regression." *Journal of the American Statistical Association*, 106: 6–20. 4, 14

Scaccia, L. and Green, P. J. (2003). "Bayesian Growth Curves Using Normal Mixtures with Nonparametric Weights." *Journal of Computational and Graphical Statistics*, 12: 308–331. 3, 18

Taddy, M. A. and Kottas, A. (2010). "A Bayesian Nonparametric Approach to Inference for Quantile Regression." *Journal of Business and Economic Statistics*, 28(3): 357–369. 3, 18

Thompson, P., Cai, Y., Moyeed, R., Reeve, D., and Stander, J. (2010). "Bayesian nonparametric quantile regression using splines." *Computational Statistics and Data Analysis*, 54: 1138–1150. 2

Tokdar, S. T. (2007). "Towards a Faster Implementation of Density Estimation with Logistic Gaussian Process Priors." *Journal of Computational and Graphical Statistics*, 16: 633–655. 4, 6, 8

Tokdar, S. T. and Ghosh, J. K. (2007). "Posterior consistency of logistic Gaussian process priors in density estimation." *Journal of Statistical Planning and Inference*, 137: 34–42. 6

Tokdar, S. T., Zhu, Y. M., and Ghosh, J. K. (2010). "Density regression with logistic Gaussian process priors and subspace projection." *Bayesian Analayis*, 5(2): 316–344. 4, 18

Trenberth, K. (2005). "Uncertainty in Hurricanes and Global Warming." *Science*, 308: 1753–1754. 4

Tsionas, E. G. (2003). "Bayesian quantile inference." *Journal of Statistical Computation and Simulation*, 73: 659–674. 2

Yu, K. and Moyeed, R. A. (2001). "Bayesian quantile regression." *Statistics & Probability Letters*, 54: 437–447. 2, 11

Zhou, K. Q. and Portnoy, S. L. (1996). "Direct use of regression quantiles to construct confidence sets in linear models." *The Annals of Statistics*, 24(1): 287–306. 2