Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach

Surya T Tokdar and Joseph B Kadane
Duke University and Carnegie Mellon University

Abstract
We consider a joint Bayesian analysis of linear conditional quantile curves of a response variable within a regression setting. For a single predictor the challenging monotonicity constraint is easily met through an interpolation of two monotone curves, modeled via logistic transformations of a Gaussian process. In analyzing time trends of north Atlantic tropical cyclone intensities, we demonstrate our joint model to offer a more robust and coherent inference than a previously published analysis obtained by fitting separate models for every quantile. By borrowing information across the entire conditional distribution, we show upward trends for all conditional quantiles – the previous analysis detected significant trends only in the upper tail. A multivariate extension is proposed by combining the full support univariate model with a linear projection of the predictors. The resulting single-index model embeds the first order linear heteroskedastic model as a special case while offering equal computational tractability. It offers an excellent fit in analyzing infant birthweight data, with the posterior concentrating entirely on models that are heteroskedastic beyond the first order.

1 Introduction
Quantile regression (Koenker and Bassett, 1978) allows one to study how an explanatory variable $X$ affects aspects of a response variable $Y$ beyond its conditional mean. Such allowance is indispensable to study the variations in the tails of the conditional distribution of $Y$ as a function of $X$. For example, in studying time trends of hurricane intensities (Elsner, Kossin, and
Jagger, 2008), the upper tail of the intensity distribution is of special interest, both because of the damage potential of the strongest hurricanes and because of their alleged connection to global warming (Trenberth, 2005). In many other environment studies, as well as in economic and social studies, the shape of the entire conditional distribution of $Y$ is suspected to change with $X$, making the conditional mean an inadequate summary of this change. In quantile regression (QR hereafter) one models the conditional quantile \( q(x, \tau) = \inf \{ q : \Pr(Y \leq q \mid X = x) \geq \tau \} \), for a suitable choice of \( \tau \in [0, 1] \), as a function of $x$ known up to a parameter vector $\beta(\tau)$. Popular QR methods (see Koenker 2005) estimate $\beta(\tau)$ by optimizing the loss \[
\sum_{i=1}^{n} \epsilon_i (\tau - I(\epsilon_i < 0)) \] where $\epsilon_i = y_i - q(x_i, \tau)$ denote the residuals for observations $(x_i, y_i)$, $1 \leq i \leq n$, on $(X, Y)$. Elegant large sample theories for such estimates have been developed for a host of model assumptions which lead to asymptotic, frequentist inference on \( q(x, \tau) \) for single or multiple choices of $\tau$ (see Koenker and Bassett, 1978; Gutenbrunner and Jurečková, 1992; Gutenbrunner, Jurečková, Koenker, and Portnoy, 1993; Koenker and Xiao, 2002; Zhou and Portnoy, 1996; Koenker and Machado, 1999; Koenker, 2005, §3.4 and the references therein).

Bayesian approaches to QR differ fundamentally from their frequentist counterparts. The former require a full specification of the distribution of $\epsilon = Y - q(X, \tau)$, which the latter avoid through large sample calculations. Yu and Moyeed (2001) and Tsionas (2003) introduced an asymmetric Laplace distribution for $\epsilon$ which has a close connection with the optimization criterion mentioned earlier. This has been subsequently improved by Kottas and Krnjajić (2009) to a flexible semiparametric distribution for $\epsilon$ (see also Thompson et al., 2010). Their proposal can also be seen as a natural extension of the semiparametric median regression models considered in Walker and Mallick (1999), Kottas and Gelfand (2001) and Hanson and Johnson (2002). Kottas and Krnjajić (2009), however, broke important grounds by allowing the error distribution to vary with $X$ – a property that Koenker (2005, §3.2.3) describes as a necessary motivation to consider QR over the traditional mean regression model.

An immediate difficulty with the Bayesian QR models mentioned above is their inability to accommodate multiple choices of $\tau$. A model on $(q(x, \tau_1), Y - q(X, \tau_1))$ for a $\tau_1 \in [0, 1]$ automatically determines a model on $(q(x, \tau_2), Y - q(X, \tau_2))$ at any other $\tau_2 \in [0, 1]$, without accommodating the intended model for the latter quantity. This inability makes these models impractical when a single choice of $\tau$, such as the median, does not present itself – as is the
case with scientific investigations that focus on a range of \( \tau \) values, such as “the tails”. A possible solution is to model the conditional distribution function of \( Y \) – a quantity that is more easily handled – and invert it to infer about the conditional quantiles (Scaccia and Green, 2003; Taddy and Kottas, 2009). Such inversion, however, leads to complicated, uninterpretable forms of \( q(\mathbf{x}, \tau) \), making this procedure less appealing when direct inference on \( q(\mathbf{x}, \tau) \) is deemed important.

The source of difficulty in modeling \( q(\mathbf{x}, \tau) \) jointly for several \( \tau \) is the constraint that \( q(\mathbf{x}, \tau) \) must be non-decreasing in \( \tau \) at each \( \mathbf{x} \in \mathcal{X} \), where \( \mathcal{X} \) is a pre-specified domain of \( X \) under consideration. Models on \( q(\mathbf{x}, \tau) \) that lend themselves to interpretability, usually violate this constraint – a phenomenon termed “Caution: Quantiles Crossing” by Koenker (2005, §2.5). For frequentist procedures, this is not taken to be a serious offense because the asymptotic theory remains valid in most cases. More often than not, the violation is not apparent from the reported estimates of the \( q(\mathbf{x}, \tau) \) curves and post-estimation fixes are also available when this is not the case (Chernozhukov, Fernandez-Va, and Galichon, 2009).

For a purely Bayesian treatment of QR, it is not an option to ignore the constraint of non-decreasing \( q(\mathbf{x}, \tau) \). Furthermore, when \( q(\mathbf{x}, \tau) \) are specified for all \( \tau \in [0, 1] \), they uniquely determine the entire conditional distribution of \( Y \) and thus place strong constraints on the distribution of \( \epsilon \). Reich et al. (2009a) took into account both these issues by expanding the first order heteroskedastic model of He (1997): \( Y = \mu + X'\gamma + (\beta_0 + X'\beta)\epsilon \), where \( x'\beta \) is positive for all \( \mathbf{x} \in \mathcal{X} \) and \( \epsilon \) has a distribution \( F_\epsilon \) does not depend \( X \). This is a special case of the popular linear model \( q(\mathbf{x}, \tau) = \beta_0(\tau) + x'\beta(\tau), \tau \in [0, 1] \), which has received considerable attention in application of QR (see Koenker and Hallock, 2001, and the references therein). Note that this model is linear in the parameter vector \( \beta(\tau) \) but not necessarily in the explanatory variables whose non-linear transforms, such as those given by smooth basis functions, can be included in \( X \). The first order heteroskddastic model, however, recovers only a small subset of the linear models where \( q(\mathbf{x}, \cdot)'s \) are linear location-scale modifications of each other. This excludes the realistic scenario where changes in \( X \) makes the conditional distribution of \( Y \) concentrate more around a certain \( \tau = \tau_1 \) while making it sparser around another \( \tau = \tau_2 \).

In this paper we explore a more flexible Bayesian specifications of the linear quantile regression model \( q(\mathbf{x}, \tau) = \beta_0(\tau) + x'\beta(\tau) \) jointly for \( \tau \in [0, 1] \). We show that for the special case where the dimension \( p \) of \( X \) equals 1, it
is relatively straightforward to construct a model that gives dense support in the set all legitimate linear specifications of \( q(x, \tau) \). We introduce and develop a construction based on transformations of smooth Gaussian processes. We present a discussion of parameter interpretation, posterior computing, posterior inference and assessment of heteroskedasticity beyond the first order. A detailed illustration is provided with a trend analysis of hurricane intensities in the north Atlantic basin based on the data used by Elsner et al. (2008). Our analysis indicates an upward time-trend of the quantiles of hurricane intensities over the last couple of decades for almost the entire range of \( \tau \in [0,1] \). Elsner et al. (2008) indicate upward movement only for the upper quantiles, with no statistically significant movement detected in the middle and lower quantiles. We demonstrate that our estimated slopes are similar to those obtained in the individual analyses, but our joint model facilitates borrowing of information across the whole range of \( \tau \) leading to a tighter inference at each point.

We next illustrate that for a vector valued predictor \( X \), construction of a valid linear \( q(x, \tau) \) involves a system of functions satisfying a stringent structural constraint. This structural constraint does not admit a mathematically tractable representation, making model specification exceedingly difficult. One possible fix is to specify a larger model that ensures linearity of \( q(x, \tau) \) and then truncate it to the valid set in which \( \tau \mapsto q(x, \tau) \) is increasing. Unfortunately, this approach is computationally infeasible when an exact likelihood computation is warranted. Dunson and Taylor (2005) propose a partial solution where a pseudo-likelihood is used based on a finite number of \( \tau \) values selected from \([a, b]\). Even this approach becomes virtually intractable when the chosen set of values is densely packed. An alternative fix is to maintain computational tractability at the expense of model support where one specifies a prior distribution that gives full mass only to a small but easily characterized subset of all valid linear specifications of \( q(x, \tau) \). A notable example of this is the first order heteroskedastic model of Reich et al. (2009a) described earlier. Along the same line, we explore a simple combination of our full-support one dimensional model with a univariate linear projection of \( X \) to identify an easily characterized subset. The resulting single-index model is shown to embed the first order heteroskedastic model as a well identified special case and is computationally more tractable. We illustrate this model with an analysis of infant birthweight data (Abrevaya, 2001) where \( X \) includes various demographic and pregnancy related attributes of the mother. The posterior distribution clearly indicates heteroskedasticity
beyond the first order. More importantly, the single-index model provides as good a fit as the unconstrained ensemble of individual fits at each $\tau$.

2 Linear Quantiles: The Univariate Case

2.1 The Model

For a univariate $X$, we can assume without loss of generality that $X$ is bounded and convex, i.e., $X$ 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 $X$ to be bounded is unavoidable for a valid linear specification of $q(x,\tau) = \beta_0(\tau) + x\beta_1(\tau)$, because the only non-intersecting lines on an unbounded $X$ are the parallel lines. Convexity of $X$ is not restrictive, because lines that do not intersect each other over a given set remain non-intersecting also over its convex hull.

A linear specification $q(x,\tau) = \beta_0(\tau) + x\beta(\tau), \tau \in [0,1]$ is monotonically increasing in $\tau$ for every $x \in X = [-1,1]$ if and only if

$$q(x,\tau) = \mu + \gamma x + \frac{1-x}{2}\eta_1(\tau) + \frac{1+x}{2}\eta_2(\tau)$$  \hspace{1cm} (1)

where $\eta_1(\tau)$ and $\eta_2(\tau)$ are monotonically increasing in $\tau \in [0,1]$. Therefore a model over $\eta = (\eta_1,\eta_2)$ induces via (1) a model over all valid, linear specifications of $q(x,\tau)$, provided it satisfies the boundary conditions:

$$q(x,0) = \underline{y}(x), q(x,1) = \overline{y}(x), \forall x \in X$$  \hspace{1cm} (2)

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 $\eta_1, \eta_2$ that are monotonically increasing and satisfy the above boundary conditions, we make use of monotonically increasing maps $\xi_1, \xi_2$ from $[0,1]$ onto itself and subject these to the following parametric transformations:

$$\eta_1 = \sigma_1\tilde{q}(\xi_1(\tau)), \quad \eta_2 = \sigma_2\tilde{q}(\xi_2(\tau))$$  \hspace{1cm} (3)

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 $y = -\infty, \overline{y} = \infty$, one can take $\tilde{q}$ to
give the conditional quantiles of a $N(\mu_0, \sigma_0)$ density or more generally, of a $t_\nu(\mu_0, \sigma_0)$ 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 $\bar{y}$ and $\tilde{y}$ are finite, we take $\tilde{q}$ to give the quantiles of a distribution supported over $[\bar{y}, \tilde{y}]$ and fix $\mu = \gamma = 0$ and $\sigma_1 = \sigma_2 = 1$. In (3), $\tilde{q}$ represents the target parametric model. Indeed, $q(x, \tau) = \tilde{\mu} + \gamma x + \tilde{\gamma} x \tilde{q}(\tau)$ – the parametric first order heteroskedastic model determined by linear location-scale changes of $\tilde{q}$ – whenever $\xi_i$'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 $\xi_i$'s.

Let $\omega(i, \tau)$ denote a zero-mean Gaussian process defined on $\{1, 2\} \times [0, 1]$, with covariance $\text{Cov}(\omega(i, \tau), \omega(i', \tau')) = \kappa^2 c_{ii'} \exp(-\lambda^2 (\tau - \tau')^2)$, 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

$$
\xi_i(\tau) = \frac{\int_0^\tau e^{\omega(i, t)} dt}{\int_0^1 e^{\omega(i, t)} dt}, \tau \in [0, 1], i = 1, 2. \tag{4}
$$

Then $\xi_1, \xi_2$ are monotonically increasing random functions that map $[0, 1]$ to $[0, 1]$. The transformation in (4), 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 $\xi_i$ in supporting all continuous or piecewise continuous, monotonically increasing bijections of $[0, 1]$ onto itself. The zero-mean property of $\omega$ implies that $\xi_i$ concentrates around the identity function – the logistic transform of the zero function.

To summarize, our specifications (1), (3) and (4) together define a valid, linear model on $q(x, \tau)$, with $\tilde{q}$ as the base quantile function, supporting any continuous or piecewise continuous specification. It is easy to see that the first order heteroskedastic model is a special case of our specification with $\rho = 1$, with further reduction to the homoskedastic model with $\sigma_1 = \sigma_2$. We complete the model by specifying

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

although the analyses reported in the subsequent chapters are fairly robust to the choice of prior on these parameters. A uniform prior on $\rho$ offers a broad range of dependence between the curves $\xi_1$ and $\xi_2$ and puts positive mass around the first order heteroskedastic case $\rho = 1$. Our prior on $\lambda$ 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 \(\kappa^{-2}\) ensures a finite variance for the marginal distribution of \(\omega(i, \tau)\).

2.2 Model Fitting

The conditional density \(f_x(y)\) of \(Y\) given \(X = x\) can be retrieved from \(q(x, \tau)\) as: \(f_x(y) = 1/\partial \partial \tau q(x, \tau x(y))\) where \(\tau x(y)\) solves \(y = q(x, \tau x(y))\) in \(\tau\). Therefore the log-likelihood function at \((\omega, \mu, \gamma, \sigma_1, \sigma_2)\) is simply computed as \(-\sum_{i=1}^{n} \log \partial \partial \tau q(x_i, \tau x_i(y_i))\) where \((x_i, y_i), i = 1, 2, \cdots, n\) are the observed data on \((X, Y)\). The solution \(\tau x(y)\) to \(q(x, \tau x(y)) - y = 0\) can be efficiently obtained by using Newton’s recursion: \(\tau x^{(k+1)}(y) = \tau x^{(k)}(y) - [q(x, \tau x^{(k)}(y)) - y]/\partial \partial \tau q(x, \tau x^{(k)}(y))\), where \(\tau x^{(0)}(y)\) is some initial guess in \([0, 1]\). Running this recursion would require repeated evaluations of \(q(x, \tau)\) and \(\partial \partial \tau q(x, \tau)\) at various values of \(\tau \in [0, 1]\). This can be carried out quite efficiently by storing \(\omega(1, \tau), \omega(2, \tau)\) on a fine grid of \(\tau\) values and using linear interpolation at values off the grid.

With the aid of an efficient algorithm to evaluate the log-likelihood function, we carry out model fitting via a Markov chain exploration of the posterior distribution of \((\omega, \lambda, \rho, \mu, \gamma, \sigma_1^2, \sigma_2^2)\) given data \((\kappa^2\) can be integrated out). Updating \((\omega, \lambda, \rho)\) according to their exact conditional posterior distribution remains challenging because of 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 \(\omega\) with a node-based approximation \(\omega^*(i, \tau) = \mathbb{E} [\omega(i, \tau) \mid W^* = \{\omega(j, \tau_k^*): j = 1, 2, k = 0, 1, \cdots, K\}]\), where \(K\) is a pre-specified order of approximation and \(\{\tau_k^* = k/K: 0 \leq k \leq K\}\) is a grid over \([0, 1]\). In our applications we use \(K = 10\). The surrogate process \(\omega^*\) can be easily evaluated at any \(\tau\) given \((W^*, \lambda, \rho)\) and these parameters are easily updated via a block-Metropolis sampler.

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
Figure 1: P-values from individual quantile regression analyses of north Atlantic tropical cyclone intensity against time. A substantial fluctuation indicates poor borrowing of information across \( \tau \), leading to difficulties in drawing a composite inference.
analysis includes fitting separate linear quantile regression models to maximum sustained wind speed ($W_{\text{maxST}}$) of tropical cyclones (including tropical storms) against their year of occurrence ($\text{Year}$) over a range of $\tau$ 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 $\tau$ 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\textsuperscript{1} for $\tau$ on the grid $\{0.01, 0.02, \ldots, 0.99\}$. A substantial fluctuation in the P-value plot is indicative of a poor borrowing of information across these separate studies and poses serious difficulty to a composite inference, even when focusing on short subintervals of $\tau$ values (e.g., $\tau \in [0.4, 0.5]$).

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 $y = 0, \bar{y} = \infty$ for the range of $W_{\text{maxST}}$, measured in nautical miles per hour (knots), and restrict $\tilde{q}$ to match the quantiles of a power-Pareto density

$$f(y) = \frac{ak}{\sigma} \left(\frac{y}{\sigma}\right)^{k-1} \left(1 + \left(\frac{y}{\sigma}\right)^{k}\right)^{-(a+1)}, \quad y > 0. \quad (6)$$

For a random variable $Y$ with (6) 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 $\bar{y}$ 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 $W_{\text{maxST}}$ only from satellite data. The two measurements are close but not identical, and some of the storms included 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 cyclones\textsuperscript{2} between 1900-1979. We proceed with the model in (3) with $\mu = 0$ and $\sigma_1^2, \sigma_2^2 \sim \text{Ga}(2, 2)$ which ensures each $\sigma_i^2$ is

\textsuperscript{1}Obtained through the rq() function of the R package quantreg, a 10000 Bootstrap sample is used to compute P-values.

\textsuperscript{2}Source: \url{http://weather.unisys.com/hurricane/atlantic/}
centered around 1, but with a wide spread. The ratio of $\sigma_1$ to $\sigma_2$ 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 $W_{\text{maxST}}$ on the vertical axis. Overlaid on the scatter plot are some of the quantile lines estimated from our joint model. Figure 3(b) shows the posterior credible intervals of the slopes $s_\tau = \frac{\partial}{\partial \tau} q(x, \tau)$. The solid lines in Figure 2(b),(c) are the two terminal conditional densities $f_{1981}(y)$ and $f_{2006}(y)$ obtained by inverting the posterior mean estimates of $q(1981, \tau)$ and $q(2006, \tau)$. For a visual comparison, we have included in these plots histograms of $W_{\text{maxST}}$ 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, the first 10,000 iterations were discarded and we saved every 90-th draw from the remaining iterations. Trace plots of $s_\tau$ (not shown here) exhibit no drifts and the autocorrelation of successive saved draws $s_\tau$ drop below 0.1 at lag 1, for all $\tau$.

Figures 2(a)-(b) clearly indicate an upward trend of $W_{\text{maxST}}$ across the entire range of $\tau \in [0, 1]$. Indeed, the posterior probabilities of $s_\tau$ being positive are estimated to be 95% or more for all $\tau$ between $[0.01, 0.99]$. To compare our inference on evidence toward positive increase with that of the individual analyses as adopted in Elsner et al. (2008), we have reproduced the P-value plot of Figure 1 in Figure 3 (a), overlaid with our estimates of $1 - p(s_\tau > 0 \mid \text{data})$. While the rise/fall patterns of the P-values are similar to those of our reported posterior probabilities, the numerical calibration of evidence is markedly different, particularly in the range $\tau \in (0.15, 0.7)$. Figure 3 (b) shows that the posterior medians of $s_\tau$ from our analysis provide a remarkably good smoothing of the estimates of $s_\tau$ obtained from the individual analyses. Therefore, the difference in inference on positivity lies entirely in calibrating the precision of these estimates – our joint model, with its ability to borrow information from the entire range of $\tau$, provides sharper estimation intervals and more robust inference across $\tau$.

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

11
Figure 3: Inference on slopes $s_\tau = \frac{\partial}{\partial \tau} q(x, \tau)$ from our joint fit and the individual fits to the cyclone intensity data. (a) P-values from individual fits for testing slope is zero and posterior probability of slope being negative from our joint fit. (b) Estimates of slope from the individual fits (thick gray line) and posterior means of slope from joint fit (thin black line).
Figure 4: Posterior summaries of our joint quantile regression analysis of maximum wind speed ($W_{maxST}$) of north Atlantic tropical cyclones against year (Year) of occurrence, with a t-density as the base. (a) Posterior mean of $q(x, \tau)$ for $\tau \in \{0.05, 0.1, 0.2, \cdots, 0.9, 0.95\}$ overlaid on data scatter. (b) Posterior median and 50% and 95% central credible intervals for slopes $s_\tau = \frac{\partial}{\partial \tau} q(x, \tau)$. (c)-(d) Terminal conditional densities $f_{1981}(y)$ and $f_{2006}(y)$ (solid lines) found by inverting posterior means of $q(1981, \tau)$ and $q(2006, \tau)$, overlaid on the histograms of $W_{maxST}$ pooled over the first and the last 10 years of study. The dashed curves are the base t-density $\tilde{f}$. 
by the $t_1(100, 8)$. This choice is not entirely suitable for $W_{\text{maxST}}$ as it gives $\bar{y} = -\infty$. 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 (3), with $\mu \sim t_1(0, 1)$, $\gamma \sim t_1(0, 1)$ and $\sigma_1^2, \sigma_2^2 \sim \text{Ga}(2, 2)$. Figure 4 shows a summary of fits with this model. The estimated quantile lines and the credible intervals on slopes 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 $s_\tau$ being positive also remain mostly unchanged (Figure 5 (a)), 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. Interestingly, the estimated posterior medians of $s_\tau$ match the individual analyses estimates even better (Figure 5(b)).

![Figure 5](attachment:image.png)

Figure 5: Inference on slopes $s_\tau = \frac{\partial}{\partial \tau} q(x, \tau)$ from our joint fit and the individual fits to the cyclone intensity data – with a t-density as the base. (a) P-values from individual fits for testing slope is zero and posterior probability of slope being negative from our joint fit. (b) Estimates of slope from the individual fits (thick gray line) and posterior means of slope from joint fit (thin black line).

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 $t(100, 8)$ base
Figure 6: Informal diagnosis for heteroskedasticity beyond the first order (see text for details). The power-Pareto base model supports quantile lines with ever increasing slopes, while the t base model shows change in sign in the rate of increase of slopes – indicating heteroskedasticity beyond linear location-scale change.

The model produces a posterior that concentrates over quantiles curves that are heteroskedastic beyond the first order. This is illustrated in Figure 6, where we plot the estimated values of $P(\Delta s_\tau > 0 \mid \text{data})$ under the t model and the power-Pareto model. Here $\Delta s_\tau$ is computed by differencing $s_\tau$ on $\tau \in \{0.01, 0.02, \ldots, 0.99\}$. A necessary condition for quantile curves to be first order heteroskedastic is that $\Delta s_\tau$ does not change its sign in the interior of $[0, 1]$. Therefore a lower bound on $P(q \text{ is heteroskedastic beyond first order} \mid \text{data})$ is given by the posterior probability of a sign change of $\Delta s_\tau$. 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 $\Delta s_\tau$ to the nearest tenth before checking for a change in sign. 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 $q(x, \tau)$ for $\tau$ close to zero.
4 Linear Quantiles: The Multivariate Case

4.1 Model

By interpreting $\eta_1$ and $\eta_2$ in (1) 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 $(1, a'_k)$ with $a_k \in \mathcal{X}$, $k = 1, \cdots, p + 1$ and let $A = [a_1 : \cdots : a_{p+1}]$. Define

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

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 $q(x, \tau)$ 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 $q(x, \tau)$ on the whole of $\mathcal{X}$, which is equivalent to verifying this at the extreme point of the convex hull of $\mathcal{X}$, takes the form of an overdetermined system whenever the number of these extreme points exceed $p + 1$. For most choices of $\eta$, the monotonicity condition fails to hold, making the model specifications in (7) computationally intractable (see, however Reich et al., 2009b, for an interesting construction).

This motivates us to look for computationally tractable alternatives to (7), 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 $\{x \in \mathbb{R}^p : \alpha'x \in (a - b, a + b)\}$ and $p_{\alpha,a,b}(x)$ denote the shifted and scaled projection $(\alpha'x - a)/b$. Define

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

(8)

where $\eta_i(\tau) = \sigma_i\tilde{q}(\xi_i(\tau))$, $i = 1, 2$ are exactly as in (3). Although (8) does not support all valid linear specifications of $q(x, \tau)$, it does embed as a special case the first order heteroskedastic model whenever $\eta_1 \equiv \eta_2$, which corresponds, one to one, to $\rho = 1$. Furthermore, the projection vector $\alpha$ offers a global, single-index summary of the relative influence of the components of $x$.

We model $\alpha$ with a $p$-variate $t$-distribution: $\alpha \mid \sigma_\alpha^2 \sim \mathcal{N}(0, \sigma_\alpha^2 I_p)$, $1/\sigma_\alpha^2 \sim \mathcal{G}(1/2, 1/2)$, 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 edged \( a,b \) are fixed as: \( a = (\max_i \alpha' x_i + \min_i \alpha' x_i)/2 \) and \( b = (\max_i \alpha' x_i - \min_i \alpha' x_i)/2 \) which ensures 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, between 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 cigarettes 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 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 (8), with \( \bar{q}(\tau) \) giving the quantiles of the \( t_1(4000, 320) \) density which gives 95% mass to (0, 8000). We take \( \mu \sim t_1(4000, 320/3) \) and the same prior is used on each component of \( \gamma \). The variance inflation parameters are modeled as \( \sigma_1^2, \sigma_2^2 \overset{iid}{\sim} \text{Ga}(2, 2) \). Although the model is fitted to the standardized versions of \( x_i \)'s, the posterior summaries reported below correspond to the actual origin and scale.

Figures 7(a)-(m) show posterior medians and central 95% credible inter-

\(^3\)Obtained from National Bureau of Economic Research: \url{www.nber.org/natality/1997/}
Figure 7: Quantile regression analysis of Birthweight data. (a)-(m) Posterior median and 95% credible intervals for slopes from the joint fit overlaid on estimated values and 95% confidence intervals from individual fits. (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.
vals for the slope parameters $s_j(\tau) = \frac{\partial}{\partial x_j} q(x, \tau)$ 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 heteroskedasticity beyond the first order, in fact, our Monte Carlo approximation to $P(q$ is heteroskedastic beyond the first order $|$ 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 same criteria applied on the mother as mentioned earlier. Figure 7(u) shows the graphs of average prediction errors $\frac{1}{5000} \sum_{i=1}^{5000} \epsilon_i^*(\tau)(\tau - I(\epsilon_i^*(\tau) < 0))$ against $\tau$, for test data residuals $\epsilon_i^*(\tau) = y_i^* - \hat{q}(x_i^*, \tau)$ where $\hat{q}(x_i^*, \tau)$ equals the estimated value of $q(x_i^*, \tau)$ for the individual fits and equals the posterior mean of $q(x_i^*, \tau)$ 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 7(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 $\tau$. 

19
5 Conclusions

The hurricane intensity analysis presented in Section 3 clearly indicates the advantages of a simultaneous quantile regression analysis over individual fits. A simultaneous analysis has the flexibility to borrow information across cases and offer tighter inference at each case. The resulting difference in inference can have nontrivial implications on an overall summary and interpretations of results. Our analysis of the infant birthweight data (Section 4) shows that for multivariate predictors, the single-index specification has enough richness to capture the complexities of real world data. It is also evident that it generalizes the first order heteroskedastic 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. The logistic Gaussian process construction presented here is attractive both from computational and theoretical perspectives. In a recent work, 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.

References


