

Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach

Surya Tokdar

Duke University, Durham NC, USA

Joseph B Kadane

Carnegie Mellon University, Pittsburgh PA, USA

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 proposed 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 $Q_Y(\tau | x) := \inf\{q : P(Y \leq q | X = x) \geq \tau\}$ 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 $Q_Y(\tau | x)$, at a given τ , specifies

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

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 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 $\{(x_i, y_i) : 1 \leq i \leq n\}$ 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 Jurečková, 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 $Y = \beta_0(\tau) + x' \beta(\tau) + \epsilon$ with ϵ having the asymmetric Laplace density $f_\epsilon(\epsilon) = \text{const} \times \exp[-\epsilon(\tau - I(\epsilon < 0))]$ (Yu and Moyeed, 2001; Tsonas, 2003). Subsequent Bayesian approaches have looked into more flexible formulations of $f_\epsilon(\epsilon)$, including non-parametric formulations with extensions to the heteroscedastic case: $f_\epsilon(\epsilon | X = x)$ (Kottas and Gelfand, 2001; Gelfand and Kottas, 2003; Kottas and Krnjajić, 2009; Thompson et al., 2010).

Arguably, this “one τ at a time” fitting of (1) does not do justice to the full potential of the model, which lies in the simultaneous description

$$\{Q_Y(\tau | x) = \beta_0(\tau) + x' \beta(\tau); \quad 0 \leq \tau \leq 1\} \quad (2)$$

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 $\tau \in \{0.01, 0.02, \dots, 0.99\}$. 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 one τ at a time approach poses an even bigger challenge to a Bayesian analysis of (1). The sampling densities for Y from two different values of τ are usually different from each

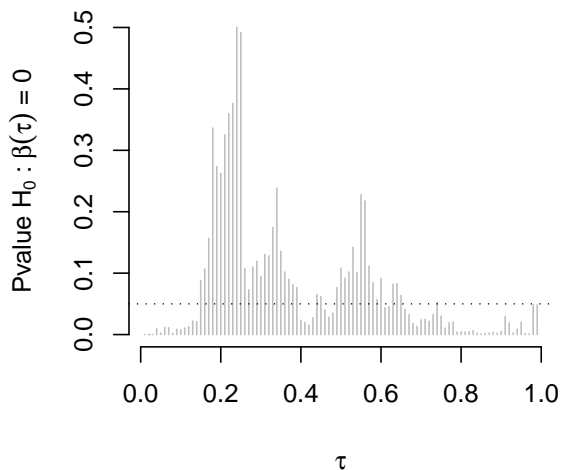


Figure 1: P-values from individual linear quantile regression analyses of north Atlantic tropical cyclone intensity against time. A substantial fluctuation indicates poor borrowing of information across τ , leading to difficulties in drawing inference.

other, resulting in an incoherent analysis when pooling is performed. The problem of simultaneous inference, too, 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 $\tau \mapsto Q_Y(\tau | x)$ 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). There have been some attempts in the literature to avoid the monotonicity problem altogether by specifying a nonparametric model for the conditional distribution $F_Y(y | x)$ and then inverting it to derive conditional quantile curves $Q_Y(\tau | x)$ (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 $\tau \mapsto Q_Y(\tau | x)$ is strictly increasing for each $x \in \mathcal{X}$, (2) uniquely determines a conditional sampling density for Y in the form:

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

where $\tau_x(y)$ solves $y = Q_Y(\tau | x)$ in τ . A likelihood function over the monotonicity preserving

choices of $(\beta_0(\cdot), \beta(\cdot))$ is then simply defined as $\prod_{i=1}^n f_Y(y_i | x_i)$.

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 obtains 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 $f_Y(y | x)$ to be linear location-scale transformations of each other. On the other hand, it is computationally much more tractable than the recent proposal in Reich et al. (2010) which offers greater model coverage.

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 $Q_Y(\tau | x) = \beta_0(\tau) + x\beta_1(\tau)$, 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 $Q_Y(\tau | x) = \beta_0(\tau) + x\beta(\tau)$, $\tau \in [0, 1]$ is monotonically increasing in τ for every $x \in \mathcal{X} = [-1, 1]$ if and only if*

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

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

Proof. If $Q_Y(\tau | x)$ 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 $Q_Y(\tau | x)$ 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 $Q_Y(\tau | x)$ of the form (2) can be expressed as (4) by taking $\mu = 0$, $\gamma = 0$, $\eta_1 = Q_Y(\cdot | -1)$ and $\eta_2 = Q_Y(\cdot | 1)$. \square \square

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

$$Q_Y(0 | x) = \underline{y}(x), Q_Y(1 | x) = \bar{y}(x), \forall x \in \mathcal{X} \quad (5)$$

where $(\underline{y}(x), \bar{y}(x))$ gives the range of Y given $X = x$. We restrict attention to the special case where this range does not change with x , $\underline{y}(x) \equiv \underline{y}$, $\bar{y}(x) \equiv \bar{y}$, 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:

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

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 $\tilde{Q}(1) = \bar{y}$. For $\underline{y} = -\infty, \bar{y} = \infty$, 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 \underline{y} and \bar{y} are finite, we take \tilde{Q} to give the quantiles of a distribution supported over $[\underline{y}, \bar{y}]$ 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 $\text{Cov}(\omega(i, \tau), \omega(i', \tau')) = \kappa^2 c_{ii'} \exp(-\lambda^2(\tau - \tau')^2)$, 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

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

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 $Q_Y(\tau | x)$, 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 homoskedastic model with $\sigma_1 = \sigma_2$. We complete the model by specifying

$$(\rho, \lambda^2, \kappa^{-2}) \sim \text{Unif}(0, 1) \times \text{Ga}(5, 1/10) \times \text{Ga}(3, 1/3), \quad (8)$$

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 $\{t_l = l\delta : l = 0, 1, \dots, L\}$. 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 $\tau = t_l$, $l = 0, 1, \dots, L$. 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

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

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

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

where $\tau_x^{(0)}(y)$ is some initial guess in $[0, 1]$. Running this recursion would require repeated evaluations of $Q_Y(\tau | x)$ and $\frac{\partial}{\partial \tau} Q_Y(\tau | x)$ 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 $\frac{\partial}{\partial \tau} Q_Y(\tau_x(y) | x)$ by $\{Q_Y(t_l | x) - Q_Y(t_{l-1} | x)\} / (t_l - t_{l-1})$.

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 $\omega^*(i, \tau) = \mathbb{E}[\omega(i, \tau) | W^* := \{\omega(j, \tau_k^*) : j = 1, 2, k = 0, 1, \dots, K\}]$, where K is a pre-specified order of approximation and $\{\tau_k^* = k/K : 0 \leq k \leq K\}$ 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 ($W_{\max ST}$) 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 tests¹ for τ on the grid $\{0.01, 0.02, \dots, 0.99\}$. As previously mentioned, a substantial fluctuation in the P-value plot is indicative of a poor borrowing of

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

information across these separate studies and poses serious difficulty to inference, even when focusing on short subintervals of τ 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 $\underline{y} = 0, \bar{y} = \infty$ 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

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

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 $y \rightarrow 0$. 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 cyclones² between 1900-1979. We proceed with the model in (6) with $\mu = 0$ and $\sigma_1^2, \sigma_2^2 \stackrel{\text{IID}}{\sim} \text{Ga}(2, 2)$ 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 3(b) shows the posterior credible intervals of the slopes $s_\tau = \frac{\partial}{\partial x} Q_Y(\tau | x)$. The solid lines in Figure 2(b),(c) are the two terminal conditional densities $f_Y(y | 1981)$ and $f_Y(y | 2006)$ obtained by inverting the posterior mean estimates of $Q_Y(\tau | 1981)$ and $Q_Y(\tau | 2006)$. 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)-(b) 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 are estimated to be 95% or more for all τ 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 $1 - p(s_\tau > 0 | \text{data})$. 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,

²Source: <http://weather.unisys.com/hurricane/atlantic/>

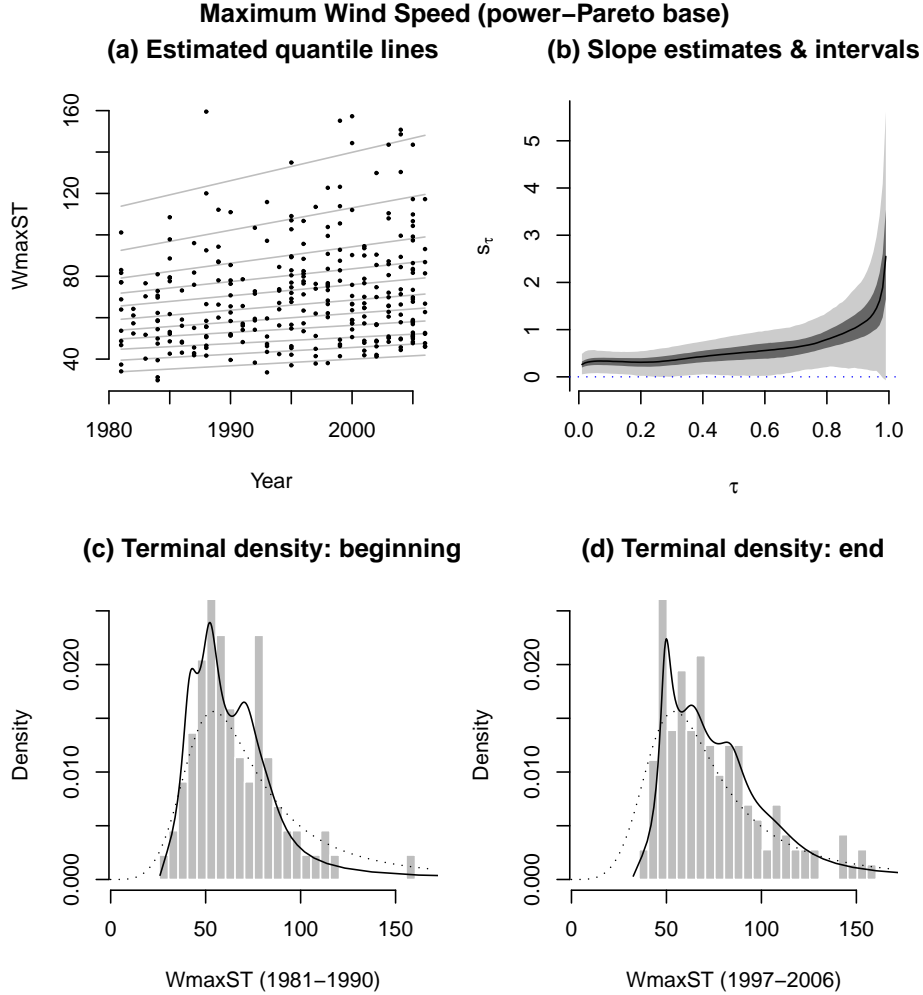


Figure 2: Posterior summaries of our joint quantile regression analysis of maximum wind speed ($W_{\max ST}$) of north Atlantic tropical cyclones against year ($Year$) of occurrence. (a) Posterior mean of $Q_Y(\tau | x)$ for $\tau \in \{0.05, 0.1, 0.2, \dots, 0.9, 0.95\}$ overlaid on data scatter. (b)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

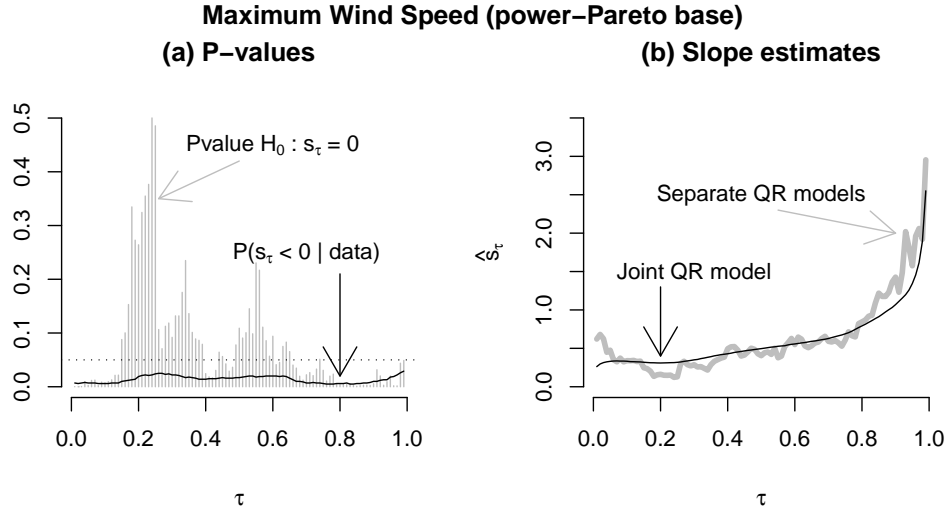


Figure 3: Inference on slopes $s_\tau = \frac{\partial}{\partial x} Q_Y(\tau | x)$ from our joint fit and the individual fits to the cyclone intensity data. (a) P-values from individual fits for testing slope is zero and posterior probability of slope being negative from our joint fit. (b) Estimates of slope from the individual fits (thick gray line) and posterior means of slope from joint fit (thin black line).

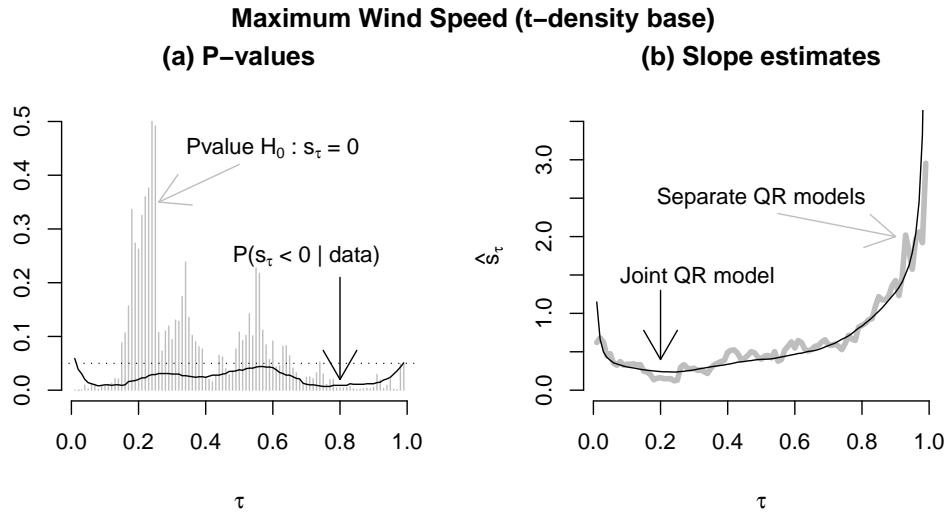


Figure 4: Inference on slopes $s_\tau = \frac{\partial}{\partial x} Q_Y(\tau | x)$ from our joint fit and the individual fits to the cyclone intensity data – with a t-density as the base. (a) P-values from individual fits for testing slope is zero and posterior probability of slope being negative from 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).

particularly in the range $\tau \in (0.15, 0.7)$. Figure 3 (b) shows that the posterior medians of s_τ from our analysis provide a remarkably good smoothing of the estimates of s_τ 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 τ , provides sharper estimation intervals and more robust inference across τ .

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 $t_1(100, 8)$. 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 $\mu \sim t_1(0, 1)$, $\gamma \sim t_1(0, 1)$ and $\sigma_1^2, \sigma_2^2 \stackrel{iid}{\sim} \text{Ga}(2, 2)$. Upon fitting this model to data, we find that the estimated quantile lines and the credible intervals on slopes (plot included in supplementary materials) 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 positive also remain mostly unchanged (Figure 4 (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_τ match the individual analyses estimates even better (Figure 4(b)).

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 model produces a posterior that concentrates over quantiles curves that are heteroscedastic beyond the first order. This is illustrated in Figure 5, 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 Δs_τ is computed by differencing s_τ on $\tau \in \{0.01, 0.02, \dots, 0.99\}$. 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 $P(Q_Y \text{ is heteroscedastic beyond first order} \mid \text{data})$ 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. 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.

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

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

Heteroscedasticity

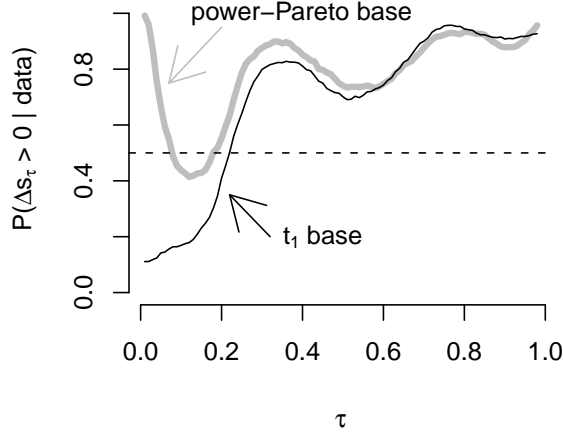


Figure 5: Diagnosis for heteroscedasticity beyond the first order (see text for details). The power-Pareto base model supports quantile lines with ever increasing slopes, while the t_1 base model shows change in sign in the rate of increase of slopes – indicating heteroscedasticity beyond linear location-scale change.

where $\eta(\tau) = (\eta_1(\tau), \dots, \eta_{p+1}(\tau))$ with each $\eta_k(\tau)$ monotonically increasing in $\tau \in [0, 1]$. It is easy to see that such $Q_Y(\tau | x)$ is monotonically increasing in $\tau \in [0, 1]$ for every x in the convex hull of $\{a_1, \dots, a_{p+1}\}$. 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 $Q_Y(\tau | x)$ 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 exceed $p + 1$. For most choices of η , the monotonicity condition fails to hold, making the model specifications in (11) computationally intractable (see, however Reich et al., 2010, for an interesting alternative construction over a finite grid of τ values).

This motivates us to look for computationally 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

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

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 $Q_Y(\tau | x)$, it does embed as a special case the first order heteroscedastic model whenever $\eta_1 \equiv \eta_2$, 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: $\alpha \mid \sigma_\alpha^2 \sim \mathbf{N}(0, \sigma_\alpha^2 I_p)$, $1/\sigma_\alpha^2 \sim \text{Ga}(1/2, 1/2)$, 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: $a = (\max_i x'_i \alpha_i + \min_i x'_i \alpha_i)/2$ and $b = (\max_i x'_i \alpha_i - \min_i x'_i \alpha_i)/2$ 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 records³ 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 (12), with $\tilde{Q}(\tau)$ giving the quantiles of the $t_1(4000, 320)$ density which gives 95% mass to $(0, 8000)$. We take $\mu \sim t_1(0, 320/3)$ and the same prior is used on each component of γ . The variance inflation parameters are modeled as $\sigma_1^2, \sigma_2^2 \stackrel{\text{iid}}{\sim} \text{Ga}(2, 2)$. Although the model is fitted to the standardized versions of x_i 's, the posterior summaries reported below correspond to the 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 heteroskedasticity beyond the first order, in fact, our Monte Carlo approximation to $P(Q_Y \text{ is heteroscedastic beyond the first order} \mid \text{data})$ is exactly 1.

While the two sets of summaries are similar in appearance (ignoring smoothness), there

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

Birthweight: Posterior Summary

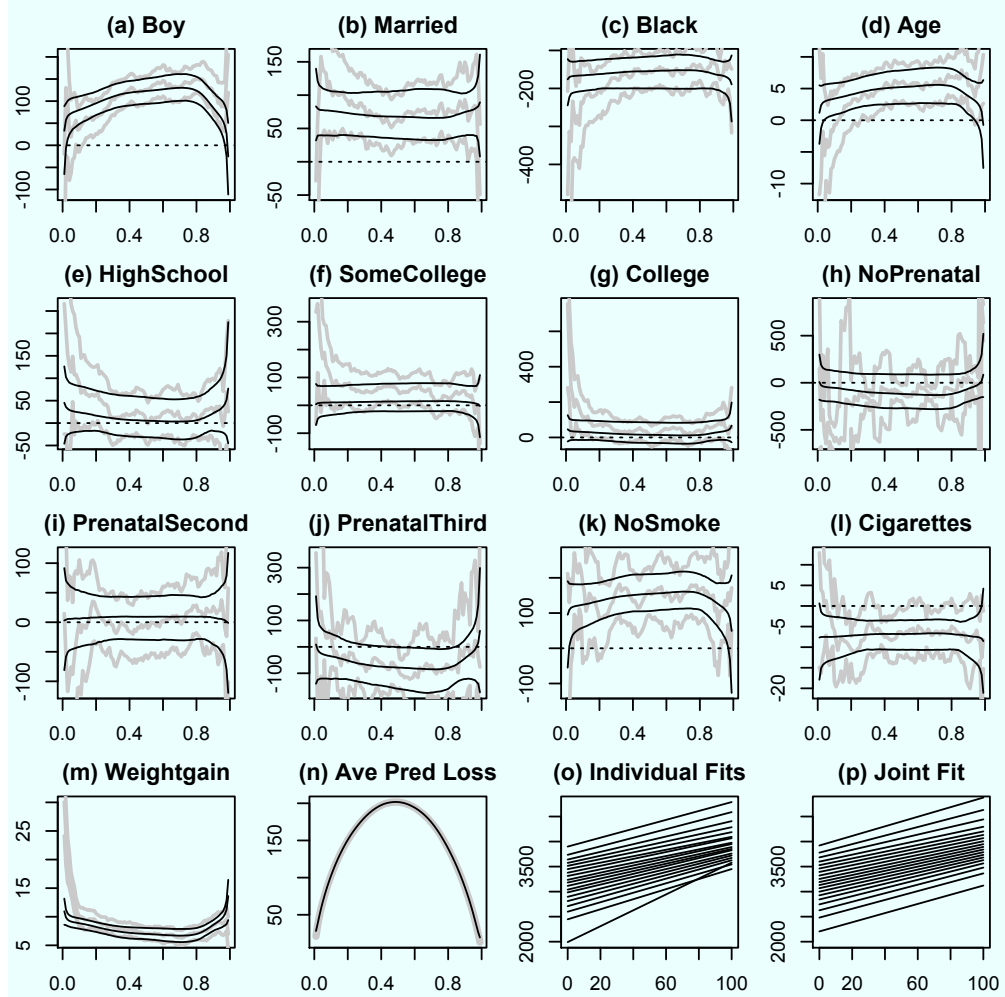


Figure 6: Quantile regression analysis of Birthweight data. (a)-(m) Posterior median and 95% credible intervals for slopes from the joint fit (thin black lines) overlaid on estimated values and 95% confidence intervals from individual fits (thick gray lines). (n) Average prediction error on test data 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 ($\tau \in \{0.05, 0.1, 0.15, \dots, 0.95\}$) from individual fits cross in the lower tail.

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 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 $\epsilon_i^*(\tau) = y_i^* - \hat{Q}_Y(\tau | x_i^*)$ where $\hat{Q}_Y(\tau | x_i^*)$ equals the estimated value of $Q_Y(\tau | x_i^*)$ for the individual fits and equals the posterior mean of $Q_Y(\tau | x_i^*)$ 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 .

5 Conclusions

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 offer tighter inference at each quantile. 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 complexities of real world data. 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.

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.

References

- Abrevaya, J. (2001). The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics* 26, 247–257.
- Banerjee, S., A. E. Gelfand, A. O. Finley, and H. Sang (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society Series B* 70(4), 825–848.
- Dunson, D. B. and J. A. Taylor (2005). Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics* 17, 385–400.
- Elsner, J. B., J. P. Kossin, and T. H. Jagger (2008). The increasing intensity of the strongest tropical cyclones. *Nature* 455, 92–95.
- Gelfand, A. E. and A. Kottas (2003). Bayesian semiparametric regression for median residual life. *Scandinavian Journal of Statistics* 30, 651–665.
- Geweke, J. and M. Keane (2007). Smoothly mixing regression. *Journal of Econometrics* 138(1), 252–290.
- Gutenbrunner, C. and J. Jurečková (1992). Regression rank-scores and regression quantiles. *The Annals of Statistics* 20, 305–330.
- Gutenbrunner, C., J. Jurečková, R. Koenker, and S. Portnoy (1993). Tests of linear hypotheses based on regression rank scores. *Journal of Nonparametric Statistics* 2, 307–333.
- He, X. (1997). Quantile curves without crossing. *The American Statistician* 51, 186–192.
- Hjort, N. D. and S. G. Walker (2009). Quantile pyramids for Bayesian nonparametrics. *The Annals of Statistics* 37(1), 105–131.
- Koenker, R. (2005). *Quantile Regression*. Cambridge U. Press.
- Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica* 46, 33–50.
- Koenker, R. and K. F. Hallock (2001). Quantile regression. *Journal of Economic Perspectives* 15, 143–156.
- Koenker, R. and J. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* 94(448), 1296–1310.
- Koenker, R. and Z. Xiao (2002). Inference on the quantile regression process. *Econometrica* 70(4), 1583–1612.
- Kottas, A. and A. E. Gelfand (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association* 96, 1458–1468.
- Kottas, A. and M. Krnjajić (2009). Bayesian nonparametric modeling in quantile regression. *Scandinavian Journal of Statistics* 36, 297–319.

- Lancaster, T. and S. J. Jun (2010). Bayesian quantile regression methods. *Journal of Applied Econometrics* 25(2), 287–307.
- Lenk, P. J. (1988). The logistic normal distribution for Bayesian, nonparametric, predictive densities. *Journal of American Statistical Association* 83, 509–516.
- Lenk, P. J. (1991). Towards a practicable Bayesian nonparametric density estimator. *Biometrika* 78, 531–543.
- Lenk, P. J. (2003). Bayesian semiparametric density estimation and model verification using a logistic gaussian process. *JCGS* 12, 548–565.
- Reich, B. J., H. D. Bondell, and H. Wang (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics* 11(2), 337–352.
- Reich, B. J., M. Fuentes, and D. Dunson (2010). Bayesian spatial quantile regression. *Journal of the American Statistical Association*, Accepted.
- Scaccia, L. and P. J. Green (2003). Bayesian growth curves using normal mixtures with nonparametric weights. *Journal of Computational and Graphical Statistics* 12, 308–331.
- Taddy, M. A. and A. Kottas (2010). A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business and Economic Statistics* 28(3), 357–369.
- Thompson, P., Y. Cai, R. Moyeed, D. Reeve, and J. Stander (2010). Bayesian nonparametric quantile regression using splines. *Computational Statistics and Data Analysis* 54, 1138–1150.
- 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.
- Tokdar, S. T. and J. K. Ghosh (2007). Posterior consistency of logistic gaussian process priors in density estimation. *Journal of Statistical Planning and Inference* 137, 34–42.
- Tokdar, S. T., Y. M. Zhu, and J. K. Ghosh (2010). Density regression with logistic gaussian process priors and subspace projection. *Bayesian Analysis* 5(2), 316–344.
- Trenberth, K. (2005). Uncertainty in hurricanes and global warming. *Science* 308, 1753–1754.
- Tsionas, E. G. (2003). Bayesian quantile inference. *Journal of Statistical Computation and Simulation* 73, 659–674.
- Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters* 54, 437–447.
- Zhou, K. Q. and S. L. Portnoy (1996). Direct use of regression quantiles to construct confidence sets in linear models. *The Annals of Statistics* 24(1), 287–306.