Joint Estimation of Quantile Planes

Surya T Tokdar, Duke University

Joint with **Yun Yang**, Florida State Unviersity
(Duke PhD 2014)

Linear regression through quantiles

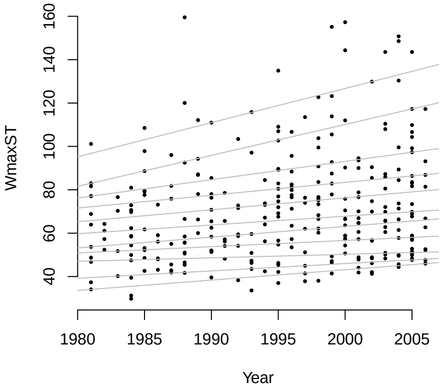
Koenker and Bassett (1978): replace $\mathbb{E}(Y|X) = \beta_0 + X^T\beta$ with

$$Q_Y(\tau \mid X) = \beta_0 + X^T\beta,$$

where

- τ is a **response proportion of interest** τ,
- $Q_Y(\tau \mid x) = \inf\{a : P(Y \leq a \mid X) \geq \tau\}$ is the corresponding (conditional) response quantile

Intensity trends of Atlantic tropical cyclones



Obvious fact: In any serious application, one looks at multiple response proportions τ

Salient features

- Analyze extreme and non-central response
  - Peer group diversity helps low achievers in elementary school!
- Capture dependence beyond changes to the mean
  - Regular vitamin intake does not improve plasma carotene level on the average, but helps the top third of the population!
- Quantify *differential* predictor effects on response distribution
  - Strongest tropical cyclones are getting stronger with time quicker than mid-range and weak cyclones!

QR as an analytic tool

- Fits are assembled from **separate single-τ analyses**
- Estimated quantiles can cross (violates probability laws)
- Little borrowing of information across response distribution

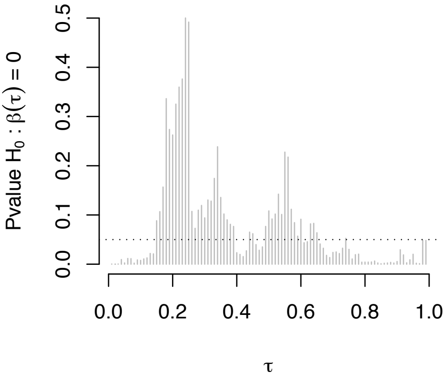
P-values from cyclone intensity analysis

Table 1 | Summary statistics

Statistic	Quantile				
	0.85	0.90	0.95	0.975	0.99
Global (2,097)					
W (m s^{-1})	51.9	55.8	62.6	68.8	75.9
Trend (m s^{-1} yr^{-1})	+0.19	+0.21	+0.18	+0.25	+0.30
s.e. (m s^{-1} yr^{-1})	0.049	0.072	0.141	0.122	0.093
P	<0.001	0.003	0.212	0.044	0.001
North Atlantic (291)					
W (m s^{-1})	48.9	54.8	60.3	72.7	77.8
Trend (m s^{-1} yr^{-1})	+0.63	+0.73	+0.81	+1.11	+1.52
s.e. (m s^{-1} yr^{-1})	0.228	0.226	0.449	0.356	NA
P	0.006	0.001	0.073	0.002	NA

From Elsner, Kossin, and Jagger (2008) published in Nature Letters

A more complete picture



In pursuit of Joint Estimation

- Model

$$Q_Y(\tau|x) = \beta_0(\tau) + x^T\beta(\tau), \quad \tau \in (0,1),$$

with function valued parameters

$$\beta_0 : (0,1) \to \mathbb{R}, \beta : (0,1) \to \mathbb{R}^p$$

- Must satisfy the monotonicity constraint

$$\beta_0(\tau_1) + x^T\beta(\tau_1) \geq \beta_0(\tau_2) + x^T\beta(\tau_2)$$

for every pair $\tau_1 > \tau_2$ and for every $x \in \mathcal{X}$ where $\mathcal{X}$ is a pre-specified domain for X

From exploratory to inference tool

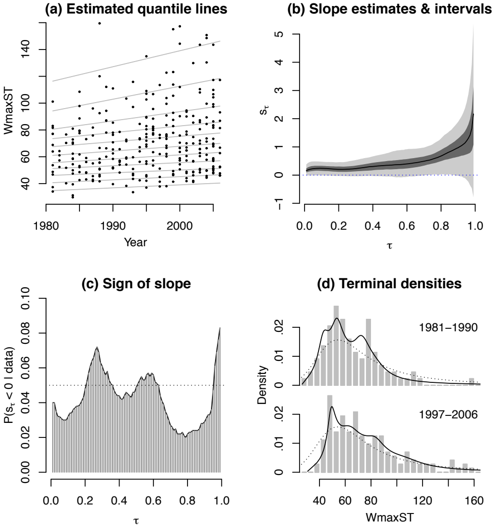
- When monotonicity holds, we have an interpretable generative model:

$$Y_i = \beta_0(U_i) + X_i^T \beta(U_i), U_i \overset{\text{IID}}{\sim} Unif(0, 1)$$

Many unsatisfactory attempts

- He (1997): imposes serious restrictions on the shape of $\beta(\tau)$
- Dunson and Taylor (2005): uses substitution likelihood, does not scale to dense τ grids
- Tokdar and Kadane (2012): complete, scalable solution when $\dim(X) = 1$; cheap shortcut when $\dim(X) > 1$

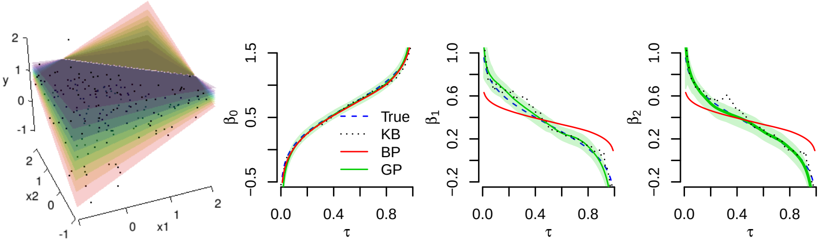
Cyclone intensity analysis from Tokdar and Kadane (2012)



Best existing approach: Reich et al. (2011)

- **Non-crossing:**
  - Bernstein basis polynomials with non-negative coefficients
  - Extending ideas from Bondell et al. (2010)
- **Prior and computing**
  - Truncated Gaussian prior distributions on coefficients
  - Gibbs sampling based Bayesian model fitting
- **BUT** require rectangular predictor domain $\mathcal{X}$
  - Getting $\mathcal{X}$ right is **important** for QR

Getting $\mathcal{X}$ right: a toy example



- X_i's simulated from the triangle $\Delta\{(-1,-1),(-1,2),(2,-1)\}$
- Crossing outside triangle but inside embedding rectangle
- GP gets it right, BP estimates are flatter

Our new approach!

- Complete characterization over **any bounded convex** $\mathcal{X}$
- **Likelihood** based estimation (penalized or Bayesian)

Characterization Theorem

- Assume non-atomic Y and 0 is an interior pt of $\mathcal{X}$
  - Non-crossing $\equiv \dot{\beta}_0(\tau) + x^T\dot{\beta}(\tau) > 0, \forall \tau \in (0,1), \forall x \in \mathcal{X}$
  - $\dot{\beta}_0(\tau) > 0 \ \forall \tau$
- Define: $a(b, \mathcal{X}) = \begin{cases} \sup_{x\in\mathcal{X}}\{-x^T b\}/\|b\| & b \neq 0, \\ \mathrm{diam}(\mathcal{X}) & b = 0. \end{cases}$
- **Theorem**. Non-crossing **if and only if**

$$\dot{\beta}_0(\tau) > 0, \quad \dot{\beta}(\tau) = \dot{\beta}_0(\tau)\frac{v(\tau)}{a(v(\tau), \mathcal{X})\sqrt{1 + \|v(\tau)\|^2}},$$

for some p-variate, real function $v(\tau) = (v_1(\tau), \cdots, v_p(\tau))^T$.

Characterization Theorem

- Assume non-atomic Y and 0 is an interior pt of $\mathcal{X}$
  - Non-crossing $\equiv \dot{\beta}_0(\tau) + x^T \dot{\beta}(\tau) > 0, \forall \tau \in (0,1), \forall x \in \mathcal{X}$
  - $\dot{\beta}_0(\tau) > 0 \ \forall \tau$
- Define: $a(b, \mathcal{X}) = \begin{cases} \sup_{x \in \mathcal{X}} \{-x^T b\} / \|b\| & b \neq 0, \\ \mathrm{diam}(\mathcal{X}) & b = 0. \end{cases}$
- **Theorem**. Non-crossing **if and only if**

$$\dot{\beta}_0(\tau) > 0, \quad \dot{\beta}(\tau) = \dot{\beta}_0(\tau) \frac{v(\tau)}{a(v(\tau), \mathcal{X}) \sqrt{1 + \|v(\tau)\|^2}},$$

for some p-variate, real function $v(\tau) = (v_1(\tau), \cdots, v_p(\tau))^T$.

Characterization Theorem

- Assume non-atomic Y and 0 is an interior pt of $\mathcal{X}$
  - Non-crossing $\equiv \dot{\beta}_0(\tau) + x^T \dot{\beta}(\tau) > 0, \forall \tau \in (0,1), \forall x \in \mathcal{X}$
  - $\dot{\beta}_0(\tau) > 0 \ \forall \tau$
- Define: $a(b, \mathcal{X}) = \begin{cases} \sup_{x \in \mathcal{X}} \{-x^T b\}/\|b\| & b \neq 0, \\ \mathrm{diam}(\mathcal{X}) & b = 0. \end{cases}$
- **Theorem**. Non-crossing **if and only if**

$$\dot{\beta}_0(\tau) > 0, \quad \dot{\beta}(\tau) = \dot{\beta}_0(\tau) \frac{v(\tau)}{a(v(\tau), \mathcal{X}) \sqrt{1 + \|v(\tau)\|^2}},$$

for some p-variate, real function $v(\tau) = (v_1(\tau), \cdots, v_p(\tau))^T$.

An almost constraint-free parametrization

- Fixed quantities:
  - A $\tau_0 \in (0,1)$, an anchoring quantile, e.g., $\tau_0 = 0.5$
  - A base pdf f_0 [e.g., $t(\nu)$], with cdf F_0, $Q_0 = F_0^{-1}$, $q_0 = \dot{Q}_0$
  - A bounded convex $\mathcal{X}$ [e.g., convex hull of observed X_i's]
- Specification

$$\beta_0(\tau_0) = \gamma_0, \beta(\tau_0) = \gamma, \qquad \beta_0(\tau) = \gamma_0 + \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} q_0(u) du$$

$$\beta(\tau) = \gamma + \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} \frac{w(u)}{a(w(u), \mathcal{X})\sqrt{1 + \|w(u)\|^2}} q_0(u) du$$

- Model parameters:
  - $\gamma_0 \in \mathbb{R}$, $\gamma \in \mathbb{R}^p$, $\sigma > 0$
  - $w : [0,1] \to \mathbb{R}^p$ (unconstrained)
  - $\zeta : [0,1] \to [0,1]$ a diffeomorphism

Relating back to the Characterization Theorem

$$\dot{\beta}_0(\tau) = \sigma q_0(\zeta(\tau))\dot{\zeta}(\tau) > 0 \text{ for all } \tau \in (0,1)$$

$$\begin{aligned}\dot{\beta}(\tau) &= \sigma \frac{w(\zeta(\tau))}{a(w(\zeta(\tau)),\mathcal{X})\sqrt{1+\|w(\zeta(\tau))\|^2}} q_0(\zeta(\tau))\dot{\zeta}(\tau) \\ &= \dot{\beta}_0(\tau)\frac{v(\tau)}{a(v(\tau),\mathcal{X})\sqrt{1+\|v(\tau)\|^2}}\end{aligned}$$

with $v := w \circ \zeta$

Role of f_0: centrally embedded linear model

- When $\zeta = \mathrm{Identity}$, $w \equiv 0$ (and $Q_0(\tau_0) = 0$)

$$Q_Y(\tau|X) = \gamma_0 + X^T\gamma + \sigma Q_0(\tau)$$

  i.e.,

$$Y = \gamma_0 + X^T\gamma + \sigma\epsilon, \quad \epsilon \sim f_0.$$

- The prior can be made to center around this linear model

Role of f_0: tail and support control

- Prior supports $f(y|x)$ with tails decaying at least as quickly as those of f_0
- To gain more control we allow a family $\{f_0(\cdot|\nu) : \nu \in S\}$ of varying tails. Our R package **qrjoint** uses the Student-t family as default.
- Can take f_0 to be supported on $[0, \infty)$ to analyze positive valued response variables

Likelihood score evaluation

- Basic math
  - $f(y|x) = \left.\frac{1}{\frac{\partial}{\partial\tau}Q_Y(\tau|x)}\right|_{\tau=\tau_x(y)}$; $\tau_x(y)$ solves $Q_Y(\tau|x) = y$
  - Log-likelihood score equals

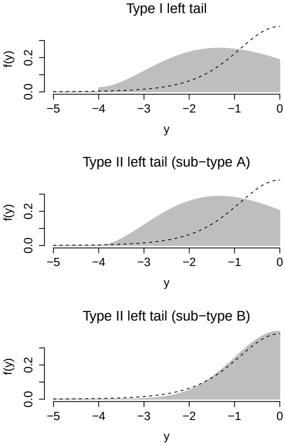
$$\sum_i \log f(y_i|x_i) = -\sum_i \log\left\{\dot{\beta}_0\left(\tau_{x_i}(y_i)\right) + x_i^T\dot{\beta}\left(\tau_{x_i}(y_i)\right)\right\}$$

- Calculations based on dense grid: $\{\tau_k = \frac{k}{N} : k = 0, \ldots, N\}$
  - Mostly quick vector-matrix multiplications
  - $a(w(\zeta(\tau)), \mathcal{X})$ is calculated as an automatic byproduct

A Bayesian implementation

- Prior specification
  - Flat/Horseshoe type shrinkage prior on γ
  - A GP prior on every w_j, $j = 1, \ldots, p$
    - Square exponential covariance function
    - Hyperpriors on range and scale
  - A logistic-GP prior on ζ:
$$\zeta(\tau) = \frac{\int_0^\tau e^{w_0(u)} du}{\int_0^1 e^{w_0(u)} du}$$
    where w_0 is a GP
- Computation:
  - Adaptive MCMC and blocking over preidctor index
  - 'Standard' GP approximations (low-rank, discretized range)
- Posterior consistency holds (well-/ill-specified wrt f_0)

Consistency holds under mild tail conditions



Defn: Type I left tail.

- $Q(0) > -\infty$
- $c(\sigma) := \lim_{t \downarrow 0} \frac{\frac{1}{\sigma} f_0(m + \frac{Q(t)-m}{\sigma})}{f(Q(t))} \in (0, \infty)$,
- $c(\sigma) \to 0$ as $\sigma \downarrow 0$.

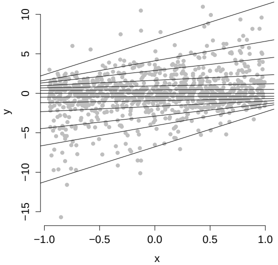
Defn: Type II left tail.

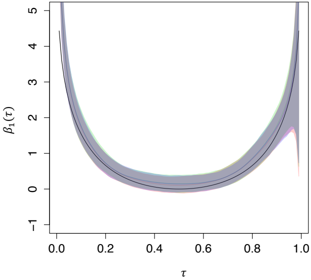
- $\lim_{t \downarrow 0} \frac{\frac{1}{\sigma} f_0(m + \frac{Q(t)-m}{\sigma})}{f(Q(t))} \to \infty$
- $u(\sigma) :=$ $\inf \left\{ t > 0 : \frac{\frac{1}{\sigma} f_0(m + \frac{Q(t)-m}{\sigma})}{f(Q(t))} \leq 1 \right\} > 0$
- $u(\sigma) \to 0$ as $\sigma \downarrow 0$.

Theorem

- **Assumptions:**
  - $\dot{\beta}^*/\dot{\beta}_0^*$ has cont. extension to $[0,1]$
  - $\exists c_0 > 0$ s.t. $\dot{\beta}_0^*(t) + x^T\dot{\beta}^*(t) \geq c_0\dot{\beta}_0^*(t)\ \forall x \in \mathcal{X}, t \in (0,1)$.
- **Result.** $f^* \in KL(\Pi)$ whenever $f_Y^*(\cdot|0)$ has type I or II tails with respect to t_ν for all small enough $\nu > 0$.

Efficient and reproducible computation

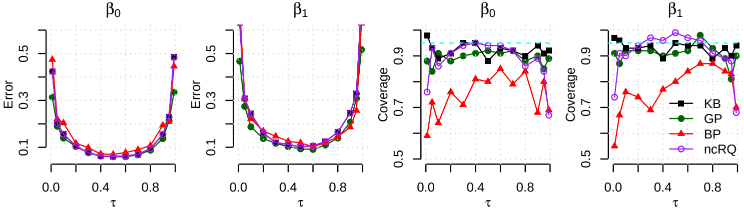




- Left: True quantile lines and synthetic data ($n = 1000$)
- Right: Posterior summaries of $\beta_1(\tau)$ from **four** runs of MCMC

Estimation accuracy

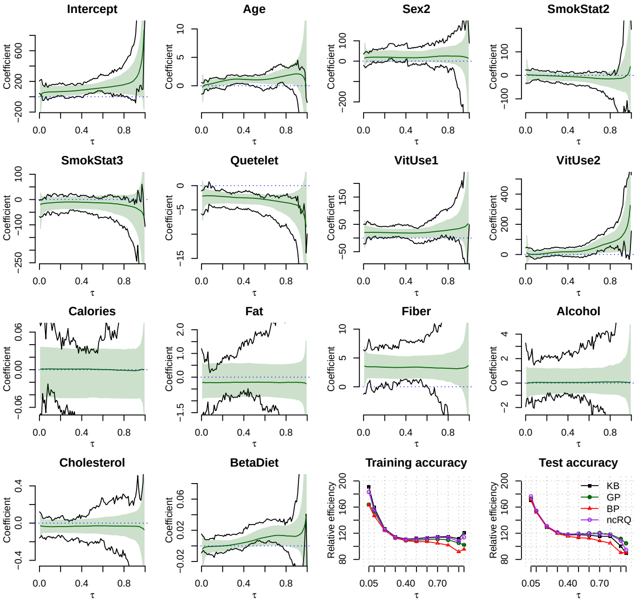
- 100 datasets from the above setting
- Pointwise mean absolute estimation error and coverage:



Application to plasma concentration of beta-carotene

- Low plasma concentration of beta-carotene may lead to higher risk of cancer
- What are the determinants of low concentration?
- $Y =$ log transform of beta-carotene concentration
- $X =$ age; smoking status; Quetelet (BMI); vitamin use, dietary intake (fat, fiber, alcohol, cholesterol, beta-carotene)

Parameter estimation



Results

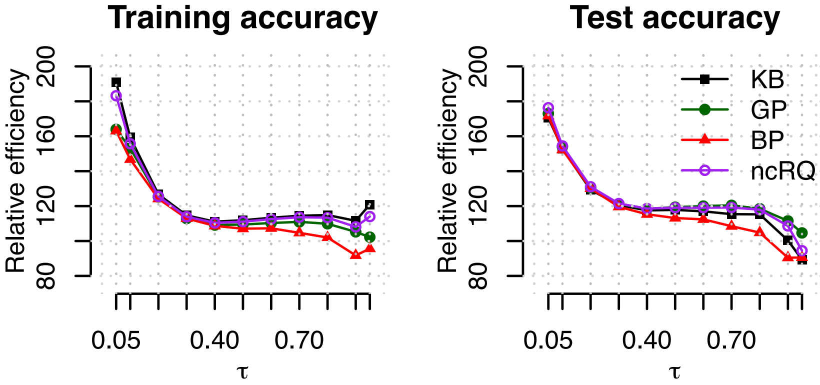
- Being female, use of vitamin and consumption of fiber have reasonably strong positive effect on plasma concentration of beta-carotene.
- Smoking and BMI have reasonably strong negative effect.
- Calories, fat, alcohol or cholesterol consumption appears to have little effect.
- Dietary intake of beta-carotene appears to have a positive effect, but the inference is not conclusive.
- More enhanced positive and negative effects, respectively for heavy vitamin use and BMI, on the upper quantiles

More dramatic effects on upper tail

j	Predictor	$\beta_j(0.9) - \beta_j(0.1)$	$\beta_j(0.9) - \beta_j(0.5)$
1	Age	$1.6_{(-0.33,4.07)}$	$0.97_{(-0.81,3.34)}$
2	Sex2	$4.6_{(-50.75,57.83)}$	$6.5_{(-42.93,57.27)}$
3	SmokStat2	$-13.43_{(-69.5,38)}$	$-5.94_{(-55.69,43.85)}$
4	SmokStat3	$-19.11_{(-94.84,31.17)}$	$-20.9_{(-90,26.57)}$
5	Quetelet	$-1.93_{(-5.61,0.91)}$	$-1.51_{(-4.79,1.02)}$
6	VitUse1	$14.37_{(-26.87,81.3)}$	$16.65_{(-22.72,81.83)}$
7	**VitUse2**	$\mathbf{113.05}_{(19.86,209.36)}$	$\mathbf{93.53}_{(12.62,185.14)}$
8	Calories	$0_{(-0.02,0.02)}$	$0_{(-0.02,0.02)}$
9	Fat	$-0.02_{(-0.37,0.36)}$	$-0.02_{(-0.37,0.32)}$
10	Fiber	$-0.16_{(-3.31,3.19)}$	$-0.22_{(-3.27,2.85)}$
11	Alcohol	$0.04_{(-0.74,1.28)}$	$0.05_{(-0.62,1.11)}$
12	Cholesterol	$0_{(-0.12,0.13)}$	$0_{(-0.12,0.12)}$
13	BetaDiet	$0.01_{(0,0.04)}$	$0.01_{(-0.01,0.03)}$

Assessment of fit

- 10 sets of random 2:1 train/test split of data
- Risk = Ave 'check' loss[1] $\rho_\tau(Y_i - \hat{\beta}_0(\tau) - X_i^T \hat{\beta}(\tau))$ at each τ
- Accuracy is inverse relative risk with LM as benchmark



[1] $\rho_\tau(r) = r\{\tau - I(r < 0)\}$

Survival analysis

- Censored observations
- Log-likelihood score calculation now changes to

$$\sum_i [(1 - c_i) \log f_Y(y_i|x_i) + c_i \log\{1 - F_Y(y_i|x_i)\}]$$
$$= \sum_i \Big[c_i \log\{1 - \tau_{x_i}(y_i)\}$$
$$- (1 - c_i) \log \left\{ \dot{\beta}_0(\tau_{x_i}(y_i)) + x_i^T \dot{\beta}(\tau_{x_i}(y_i)) \right\} \Big],$$

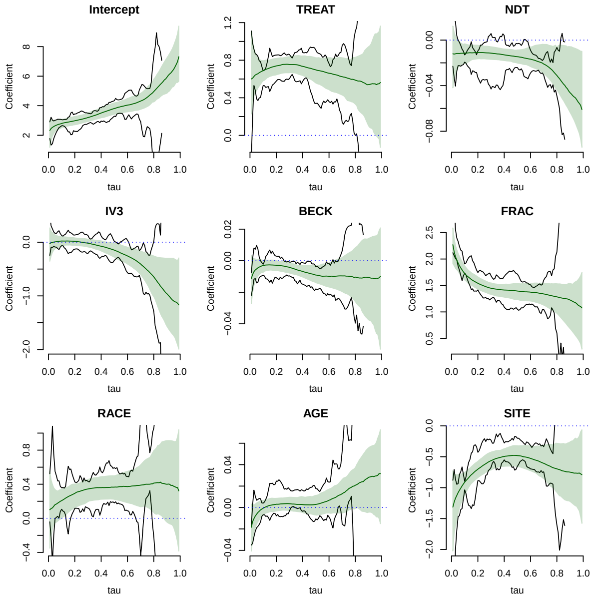
c_i = censoring status (1= right censored, 0 = observed).

- That's all that needs changing!

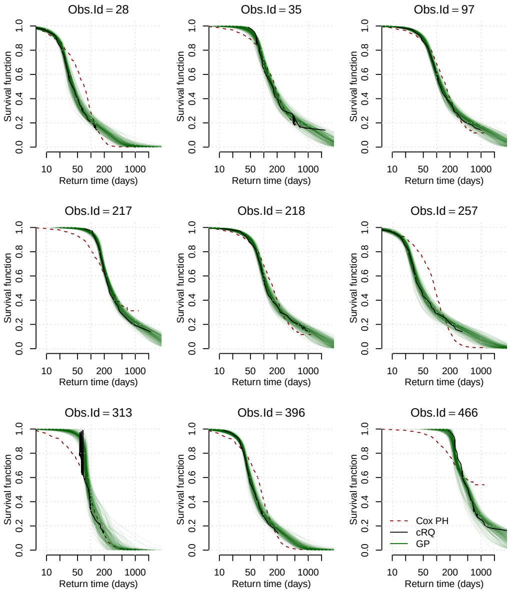
Return to drug study

- Data from University of Massachusetts Aids Research Unit IMPACT Study data (UIS, Hosmer and Lemeshow, 1998, Table 1.3)
- Response = (log) time to return to drug
- Predictors include current treatment assignment , drug use history, compliance factor, depression score, race, age, treatment site
- Total 575 subjects. Return times were right censored for 111

Parameter estimation



Survival curves – crossing!



Reference

Bondell, H. D., B. J. Reich, and H. Wang (2010). Noncrossing quantile regression curve estimation. Biometrika 97(4), 825–838.

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

Elsner, J. B., J. P. Kossin, and T. H. Jagger (2008). The increasing intensity of the strongest tropical cyclones. Nature 455(7209), 92–95.

He, X. (1997). Quantile curves without crossing. The American Statistician 51(2), 186–192.

Hosmer, D. W. and S. Lemeshow (1998). Applied Survival Analysis: Regression Modeling of Time to Event Data. New York, NY: John Wiley and Sons Inc.

Koenker, R. and G. Bassett (1978). Regression quantiles. Econometrica: Journal of the Econometric Society 46(1), 33–50.

Reich, B. J., M. Fuentes, and D. B. Dunson (2011). Bayesian spatial quantile regression. Journal of the American Statistical Association 106(493), 6–20.

Tokdar, S. T. and J. B. Kadane (2012). Simultaneous linear quantile regression: a semiparametric Bayesian approach. Bayesian Analysis 7(1), 51–72.

Package 'qrjoint'

December 15, 2016

Version 1.0-4

Date 2016-05-10

Title Joint Estimation in Linear Quantile Regression

Author Surya Tokdar <tokdar@stat.duke.edu>

Maintainer Surya Tokdar <tokdar@stat.duke.edu>

Depends R (>= 2.6), stats, graphics, grDevices, quantreg

Imports splines, coda, Matrix, kernlab

Description
Joint estimation of quantile specific intercept and slope parameters in a linear regression setting.

License GPL-2

NeedsCompilation yes

R topics documented: