Noncrossing structured additive multiple-output Bayesian quantile regression models

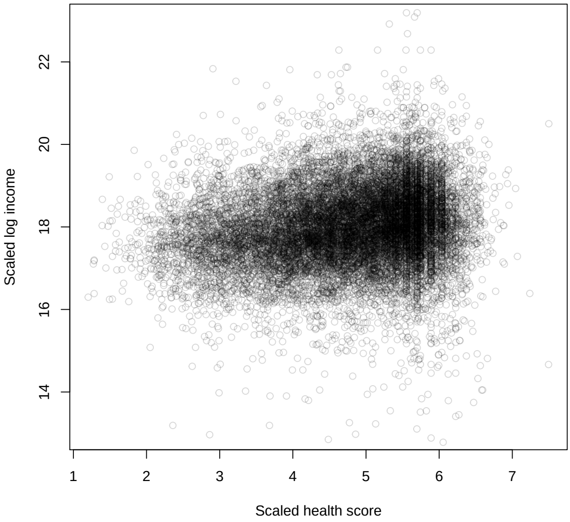
Bruno Santos

Joint work with Thomas Kneib

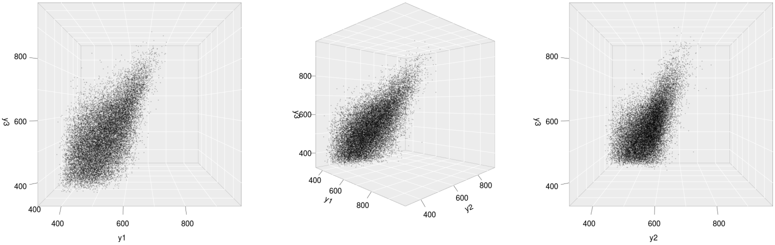
Universidade Federal do Espírito Santo
Georg-August-Universität Göttingen

3º EPEI, São Paulo

How one builds quantile regression models for two dimensional response variables using structured additive predictors?



How one builds quantile regression models for three dimensional response variables using structured additive predictors?



Quantile definition

Definition

Let X be a random variable with cumulative distribution function (cdf)

$$F(x) = P(X \leqslant x).$$

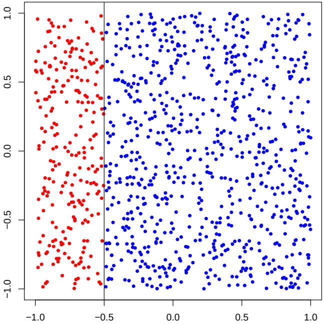
Then using the inverse of the cdf at τ, we can define

$$F^{-1}(\tau) = \inf\{x : F(x) \geqslant \tau\}$$

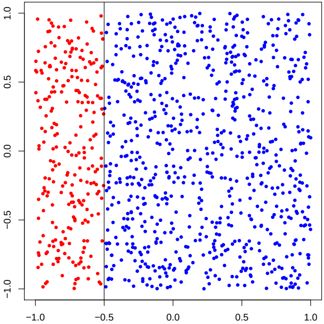
as the τth quantile of X.

- How one can define a multivariate quantile?

Tukey depth (Halfspace depth)



Tukey depth (Halfspace depth)



- Tukey depth:

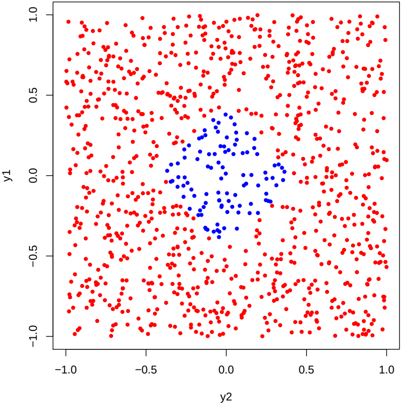
$$HD(z, P) := \inf\{P(H) : H \text{ is a closed halfspace containing } z\}.$$

Depth region

- Depth region: $D(\tau) := \{z \in \mathbb{R}^k : HD(z, P) \geq \tau\}$.

Depth region

- Depth region: $D(\tau) := \{z \in \mathbb{R}^k : HD(z, P) \geq \tau\}$.



Quantile regression models

- Quantile regression:
  - ◇ models the conditional quantiles of a response variable;
  - ◇ there is no need for a probability distribution assumption;
  - ◇ estimation by linear programming algorithms;
  - ◇ inference through asymptotic results or bootstrap, for instance;
- Bayesian quantile regression:
  - ◇ assumes an asymmetric Laplace distribution for the likelihood;
  - ◇ posterior consistent in the case of misspecified models;

Both approaches have been:

- used in a comprehensive list of areas of scientific inquiry;
- extended to a variety of different modelling approaches.

Example

$$Q_Y(\tau|X) = X\beta(\tau)$$

Example

$$Q_Y(\tau|X) = X\beta(\tau)$$

nature

Vol 455|4 September 2008|**doi:10.1038/nature07234**

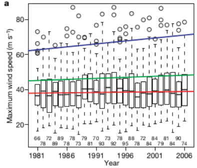
LETTERS

The increasing intensity of the strongest tropical cyclones

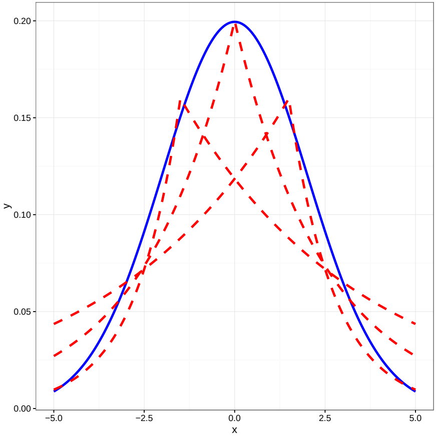
James B. Elsner[1], James P. Kossin[2] & Thomas H. Jagger[1]

Atlantic tropical cyclones are getting stronger on average, with a 30-year trend that has been related to an increase in ocean temperatures over the Atlantic Ocean and elsewhere[1–4]. Over the rest of the tropics, however, possible trends in tropical cyclone intensity are less obvious, owing to the unreliability and incompleteness of the observational record and to a restricted focus, in previous trend analyses, on changes in average intensity. Here we overcome these two limitations by examining trends in the upper quantiles of per-cyclone maximum wind speeds (that is, the maximum intensities that cyclones achieve during their lifetimes), estimated from homogeneous data derived from an archive of satellite records. We find significant upward trends for wind speed quantiles above the 70th percentile, with trends as high as 0.3 ± 0.09 m s^{-1} yr^{-1} (s.e.) for the strongest cyclones. We note separate upward trends in the estimated lifetime-maximum wind speeds of the very strongest tropical cyclones (99th percentile) over each ocean basin, with the largest increase at this quantile occurring over the North Atlantic, although not all basins show statistically significant increases. Our results are qualitatively consistent with the hypothesis that as the seas warm, the ocean has more energy to convert to tropical cyclone wind.

To quantify and determine the significance of these trends, we use quantile regression. Quantile regression as employed here is a method to estimate the change (trend) in lifetime-maximum wind speed quantile as a function of year. A quantile is a point taken from



Bayesian quantile regression - idea



Sriram et al. (2013) proved that this is posterior consistent to estimate the conditional quantiles.

Directional quantile regression for multivariate data
Hallin et al. (2010)

- Response variable is defined as $\mathbf{Y} \in \mathbb{R}^k$.
- Directional index can be defined by $\boldsymbol{\tau} \in \mathcal{B}^k := \{\mathbf{v} \in \mathbb{R}^k : 0 < ||\mathbf{v}||_2 < 1.\}$.
  - ◇ Direction: $\mathbf{u} \in \mathcal{S}^{k-1} := \{z \in \mathbb{R}^k : ||z|| = 1\}$;
  - ◇ Magnitude: $\tau \in (0, 1)$.
- Define $\boldsymbol{\Gamma}_u$, an arbitrary $k \times (k-1)$ matrix of unit vectors.
  - ◇ $(u \vdots \Gamma_u)$ is an orthonormal basis of $\mathbb{R}^k$.

Directional quantile regression for multivariate data
Hallin et al. (2010)

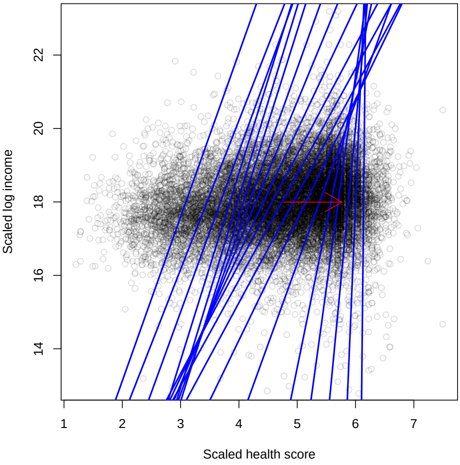
- Response variable is defined as $\mathbf{Y} \in \mathbb{R}^k$.
- Directional index can be defined by $\boldsymbol{\tau} \in \mathcal{B}^k := \{\mathbf{v} \in \mathbb{R}^k : 0 < ||\mathbf{v}||_2 < 1.\}$.
  - ◇ Direction: $\mathbf{u} \in \mathcal{S}^{k-1} := \{z \in \mathbb{R}^k : ||z|| = 1\}$;
  - ◇ Magnitude: $\tau \in (0, 1)$.
- Define $\boldsymbol{\Gamma}_u$, an arbitrary $k \times (k-1)$ matrix of unit vectors.
  - ◇ $(u \vdots \Gamma_u)$ is an orthonormal basis of $\mathbb{R}^k$.

Definition

The $\boldsymbol{\tau}$th quantile of $\mathbf{Y}$ is the τth quantile hyperplane obtained from the regression of $\mathbf{Y}_u := \mathbf{u}'\mathbf{Y}$ on the marginals of $\mathbf{Y}^{\perp} := \boldsymbol{\Gamma}_u{}'\mathbf{Y}$ with an intercept term.

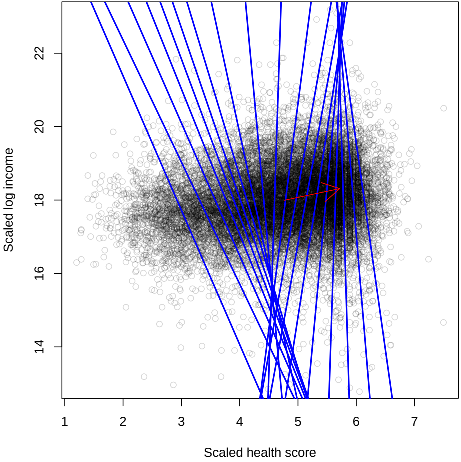
Example

Direction $\mathbf{u} = (1, 0)$



Example

Direction $\mathbf{u} = (3/\sqrt{10}, 1\sqrt{10})$



Upper and lower halfspaces

The $\boldsymbol{\tau}$th quantile of $\mathbf{Y}$ is any element of the collection Λ_τ of hyperplanes $\lambda_\tau := \{\mathbf{y} \in \mathbb{R}^k : \mathbf{u}'\mathbf{y} = \hat{\mathbf{b}}_\tau \boldsymbol{\Gamma}'_u \mathbf{y} + \hat{a}_\tau\}$, such that $(\hat{a}_\tau, \hat{\mathbf{b}}_\tau)$ are the solutions of the minimization problem

$$\min_{(a_\tau, \mathbf{b}_\tau) \in \mathbb{R}^k} E[\rho_\tau(\mathbf{u}'\mathbf{y} - \mathbf{b}_\tau \boldsymbol{\Gamma}'_u \mathbf{y} - a_\tau)].$$

Upper and lower halfspaces

The $\boldsymbol{\tau}$th quantile of $\mathbf{Y}$ is any element of the collection Λ_τ of hyperplanes $\lambda_\tau := \{\mathbf{y} \in \mathbb{R}^k : \mathbf{u}'\mathbf{y} = \hat{\mathbf{b}}_\tau \boldsymbol{\Gamma}'_u \mathbf{y} + \hat{a}_\tau\}$, such that $(\hat{a}_\tau, \hat{\mathbf{b}}_\tau)$ are the solutions of the minimization problem

$$\min_{(a_\tau, \mathbf{b}_\tau) \in \mathbb{R}^k} E[\rho_\tau(\mathbf{u}'\mathbf{y} - \mathbf{b}_\tau \boldsymbol{\Gamma}'_u \mathbf{y} - a_\tau)].$$

With predictor variables, we have

$$\lambda_\tau(\mathbf{X}) = \{\mathbf{u}'\mathbf{y} = \hat{\mathbf{b}}_\tau \boldsymbol{\Gamma}'_u \mathbf{y} + \mathbf{x}'\hat{\boldsymbol{\beta}}_\tau + \hat{a}_\tau\},$$

We can say that each element $(\hat{a}_\tau, \hat{\mathbf{b}}_\tau, \hat{\boldsymbol{\beta}}_\tau)$ define an upper closed quantile halfspace

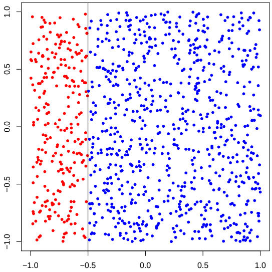
$$H^+_{\tau\mathbf{u}} = H^+_{\tau\mathbf{u}}(\hat{a}_\tau, \hat{\mathbf{b}}_\tau, \hat{\boldsymbol{\beta}}_\tau) = \{\mathbf{y} \in \mathbb{R}^k : \mathbf{u}'\mathbf{y} \geq \hat{\mathbf{b}}_\tau \boldsymbol{\Gamma}'_u \mathbf{y} + \mathbf{x}'\hat{\boldsymbol{\beta}}_\tau + \hat{a}_\tau\} \quad (1)$$

and an analogous lower open quantile halfspace switching $\geq$ for $<$.

Properties

- Probabilistic nature of quantiles:

$$P(\mathbf{Y} \in H^{-}_{\tau\mathbf{u}}) - \tau = 0,$$

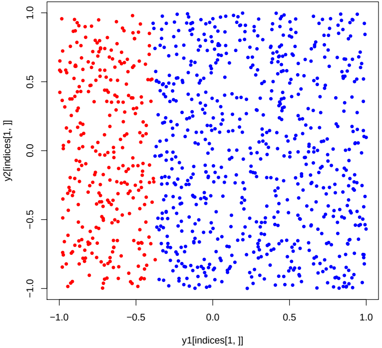


- Line joining the two probability mass centers is parallel to **u**.

Quantile region

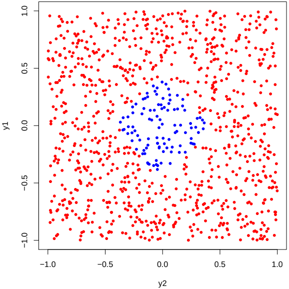
Moreover, fixing τ we are able to define the τ quantile region $R(\tau)$ as

$$R(\tau) = \bigcap_{\mathbf{u} \in \mathcal{S}^{k-1}} H^{+}_{\tau\mathbf{u}}.$$



Tukey depth and directional quantile regression output

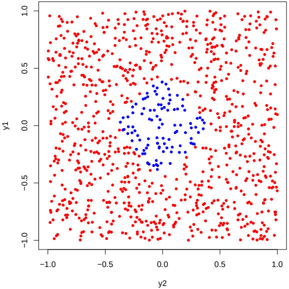
- Depth region: $D(\tau) := \{z \in \mathbb{R}^k : HD(z, P) \geq \tau\}$.
- Quantile region $R(\tau)$:



Tukey depth and directional quantile regression output

- Depth region: $D(\tau) := \{z \in \mathbb{R}^k : HD(z, P) \geq \tau\}$.

- Quantile region $R(\tau)$:



$$R(\tau) = D(\tau)$$

Bayesian directional quantile regression model
Guggisberg (2017)

Consider the mixture representation of the asymmetric Laplace distribution

$$
\begin{gathered}
Y_i|w_i \sim N(\mu + \theta w_i, \psi^2 \sigma w_i) \\
w_i \sim \text{Exp}(\sigma) \\
\Updownarrow \\
Y \sim AL(\mu, \sigma, \tau)
\end{gathered}
$$

Then one can consider that, for each direction u,

$$Y_u|\mathbf{b}_\tau, \boldsymbol{\beta}_\tau, \sigma, w \sim N(Y^\perp b_\tau + \mathbf{x}'\boldsymbol{\beta}_\tau + \theta w_i, \psi^2 \sigma w_i),$$

Bayesian directional quantile regression model
Guggisberg (2017)

Consider the mixture representation of the asymmetric Laplace distribution

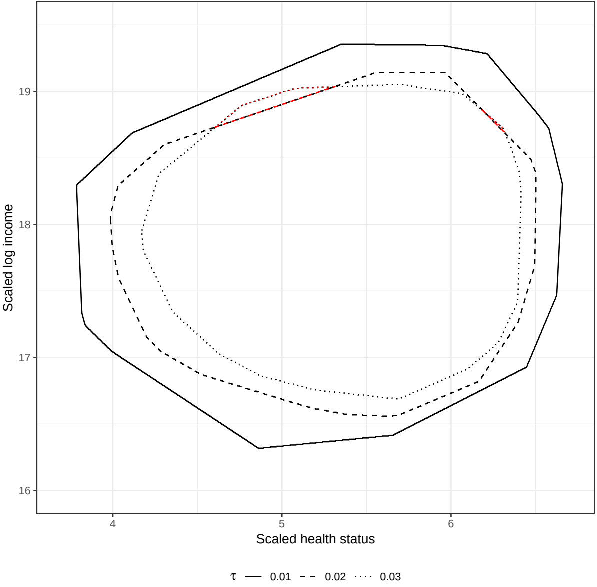
$$
\begin{gathered}
Y_i | w_i \sim N(\mu + \theta w_i, \psi^2 \sigma w_i) \\
w_i \sim \text{Exp}(\sigma) \\
\Updownarrow \\
Y \sim AL(\mu, \sigma, \tau)
\end{gathered}
$$

Then one can consider that, for each direction u,

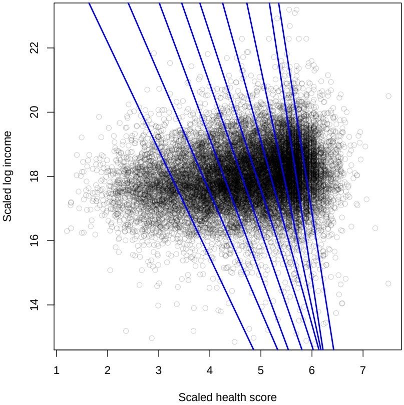
$$Y_u | \mathbf{b}_\tau, \boldsymbol{\beta}_\tau, \sigma, w \sim N(Y^{\perp} b_\tau + \mathbf{x}' \boldsymbol{\beta}_\tau + \theta w_i, \psi^2 \sigma w_i),$$

- That result makes it possible to use interesting developments of the univariate to the multivariate case.

Multiple-output crossing problem?



Possible solution



Solution: Gaussian process regression adjustment

Gaussian process regression adjusment

Rodrigues and Fan (2017)

Take the quantile function of an asymmetric Laplace distribution:

$$
\begin{aligned}
Q_{Y_u}(p|\mu,\sigma,\tau) &= F^{-1}(p;\mu,\sigma,\tau) \\
&= \begin{cases} \mu + \dfrac{\sigma}{1-\tau}\log\left(\dfrac{p}{\tau}\right), & \text{if } 0 \le p \le \tau \\ \mu - \dfrac{\sigma}{\tau}\log\left(\dfrac{1-p}{1-\tau}\right), & \text{if } \tau \le p \le 1 \end{cases},
\end{aligned}
$$

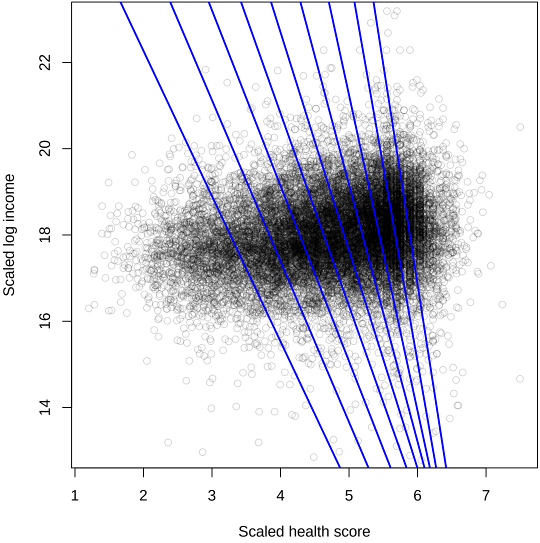
Then one assumes a Gaussian process to correlate the different τ's

$$
\begin{aligned}
\hat{Q}_s(p|\theta) &= g(\tau) + \epsilon, \quad \text{with } g(\tau) \sim GP(0, K), \\
\epsilon &\sim N(0, \Sigma),
\end{aligned}
$$

with correlation matrix given by

$$
k(\tau, \tau') = \sigma_k^2 \exp\left\{-\frac{1}{2b^2}(\tau - \tau')^2\right\},
$$

Correcting the crossing issue for each direction



Structured additive predictors
Waldmann et al. (2013)

We can also consider predictors with

- nonlinear effects;
- spatial effects;
- random effects;

We can write

$$Y_u|\mathbf{b}_\tau, \boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau, \sigma, w \sim N(\eta + \theta w_i, \psi^2 \sigma w_i),$$

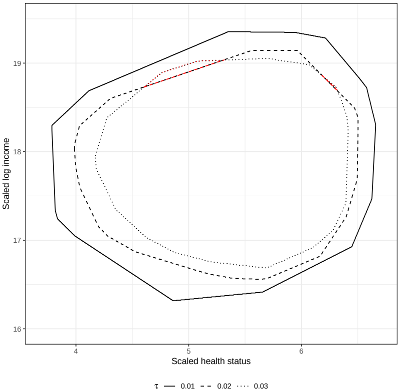
where

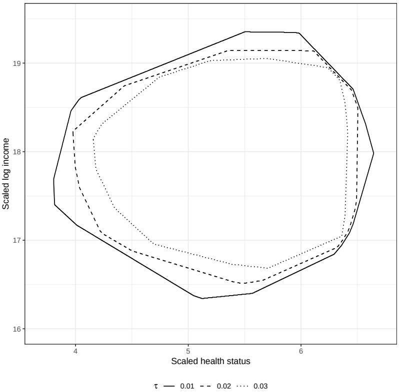
$$\eta = Y_u^\perp \mathbf{b}_\tau + X' \beta_\tau + Z' \boldsymbol{\gamma}_\tau.$$

Application - Income and Health

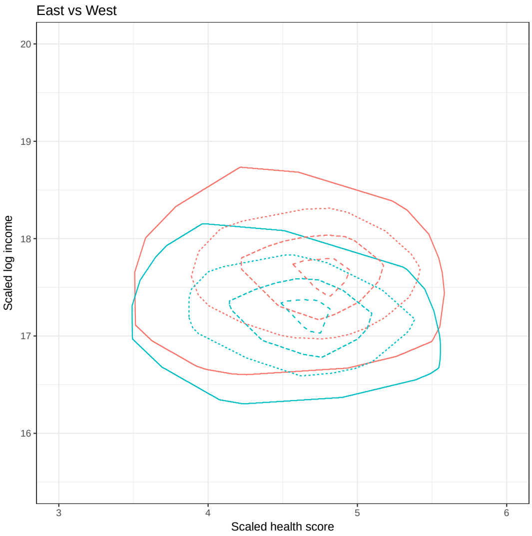
- Socio Economic Panel collected in Germany in 2012.
- 16,719 observations.
- Two dimensions of inequality in the population:
  - ◇ health and income.
- Predictor variables:
  - ◇ age
  - ◇ education
  - ◇ family status
  - ◇ Region

Crossing in quantile contours

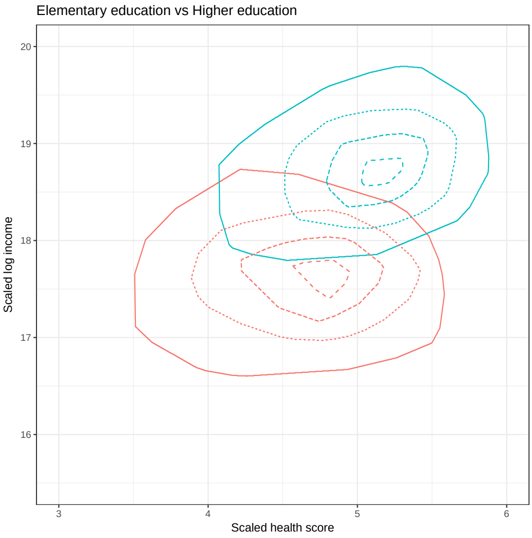




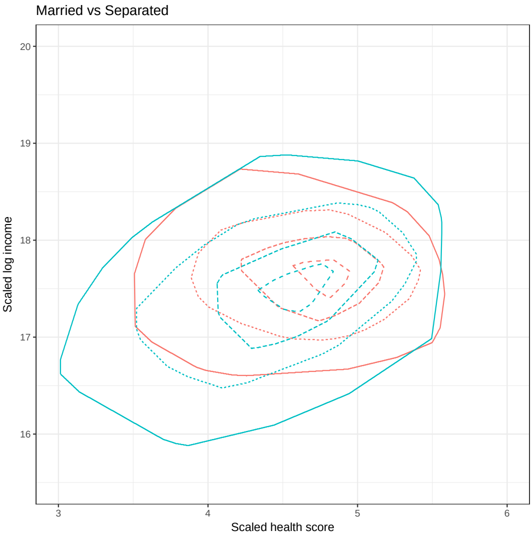
Results - Quantile contours



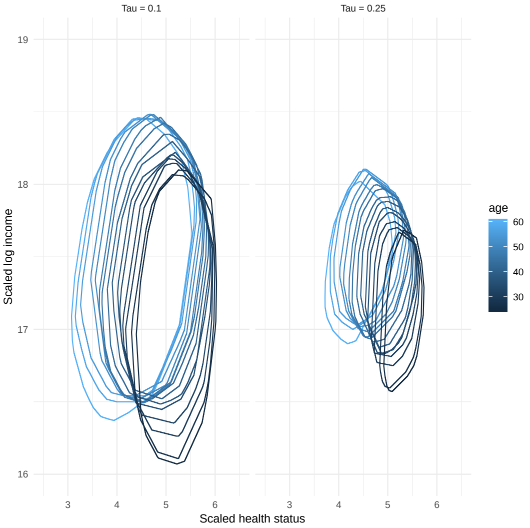
Results - Quantile contours



Results - Quantile contours



Nonlinear effect of age on the quantile contours

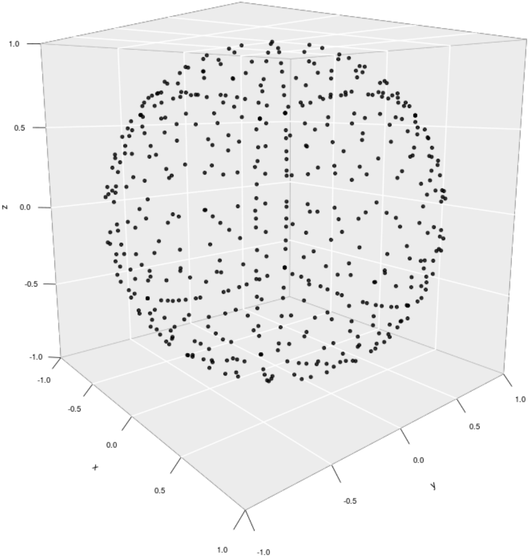


Application - ENEM

Data available online about exam students take to enter universities.

- There are more than 7 million observations in the data.
- We consider around 20.000 observations from the state of SP.
- Response variable:
  - y1: Score in natural sciences, such as Chemistry, Physics and Biology.
  - y2: Score in human sciences, such as History, Geography, Philosophy and Sociology.
  - y3: Score in Mathematics.
- Covariates available:
  - scores for different disciplines;
  - gender;
  - private and public schools;
  - Income and education of the parents;
  - ...

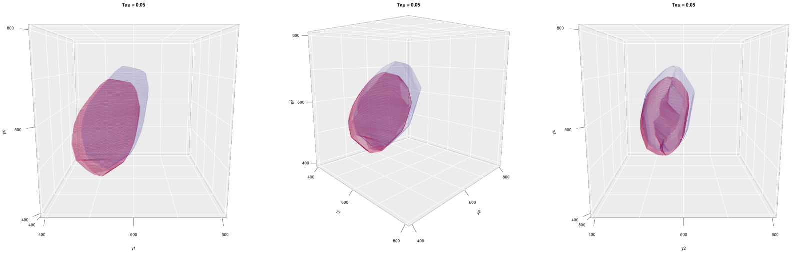
Directions



Results

Effect of private vs public schools

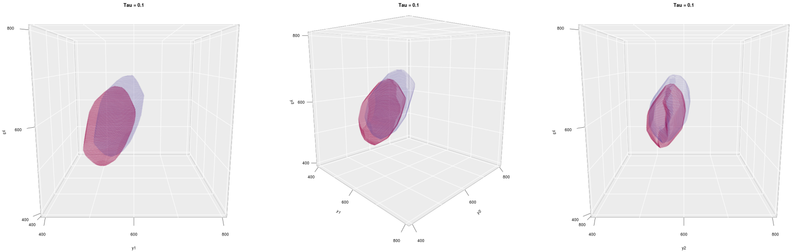
- $\tau = 0.05$



Results

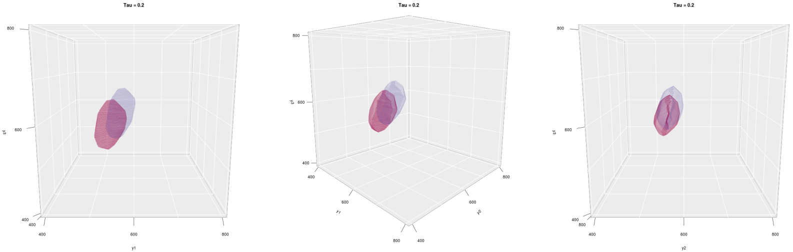
Effect of private vs public schools

- $\tau = 0.10$



Results
Effect of private vs public schools

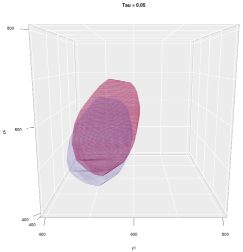
- $\tau = 0.20$

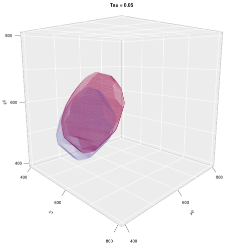


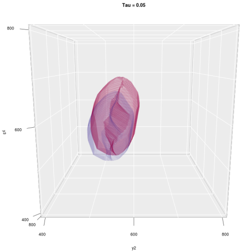
Results

Effect of men vs women

- $\tau = 0.05$



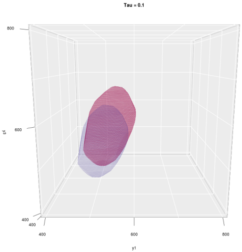


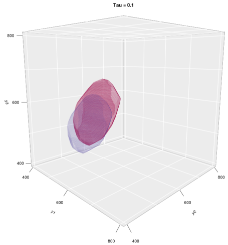


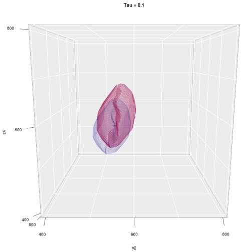
Results

Effect of men vs women

- $\tau = 0.10$



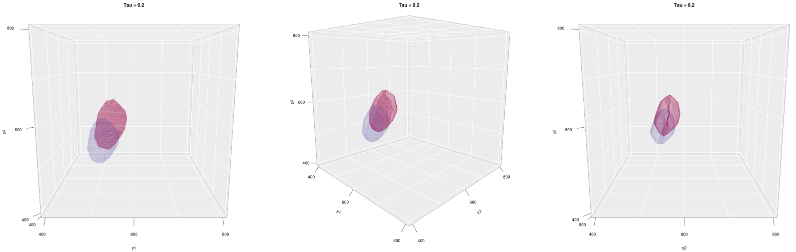




Results

Effect of men vs women

- $\tau = 0.20$



Final remarks

- Directional quantile regression models for multivariate data
  - ◇ connection to the Tukey depth concept;
  - ◇ interesting tool to study this type of response variable.
- Structured additive predictors:
  - ◇ more flexibility to the modelling process;
- Illustration on income and health inequalities:
  - ◇ conclusions depend on the interaction between direction and covariates.
  - ◇ sheds new light on how to look to the joint distribution of these variables.
- Illustration on ENEM:
  - ◇ this method is able to give a more complete picture of the effect of covariates on the different subjects considering their correlation.

Next steps

- Models for more than 3 dimensions:
  - how to define the directions?
  - how to visualize the results?
  - multivariate growth charts?

References

Guggisberg (2017).
A Bayesian Approach to Multiple-Output Quantile Regression.
Preprint on webpage at
http://www.michaelguggisberg.com/research.html.

Hallin, Paindaveine and Šiman (2010).
Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth.
In *Annals of Statistics*.

Rodrigues and Fan (2017).
Regression Adjustment for Noncrossing Bayesian Quantile Regression.
In *Journal of Computational and Graphical Statistics*.

Waldmann, Kneib, Yue, Lang and Flexeder (2013).
Bayesian semiparametric additive quantile regression.
In *Statistical Modelling*.

Thank you for the attention!