# Quantile regression: a classical and Bayesian approach

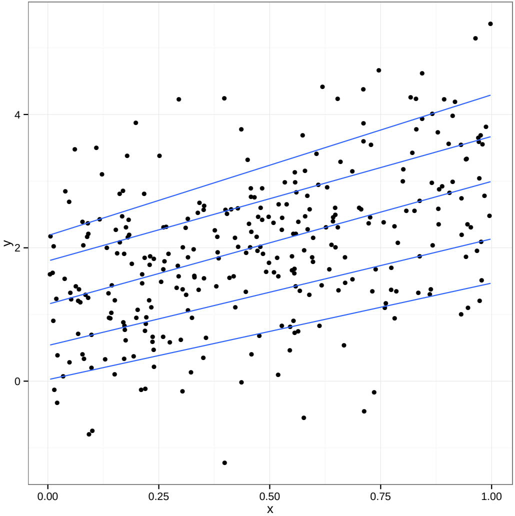
Bruno Santos

Instituto de Matemática e Estatística
Universidade Federal da Bahia

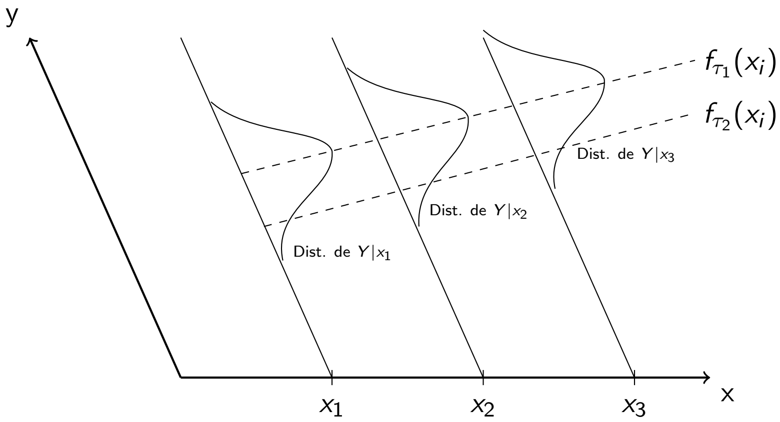
Universidad de Antofagasta
Antofagasta, Chile

# Outline

# Quantile regression models



# Quantile regression




$$
\begin{aligned}
Q_Y[\tau_j|x_i] &= f_{\tau_j}(x_i) \\
&= \beta_0(\tau_j) + \beta_1(\tau_j)x_i
\end{aligned}
$$

## Asymmetric Laplace distribution

Koenker and Machado (1999):

- Likelihood ratio test for quantile regression

If $Y \sim AL(\mu, \sigma, \tau)$, then its density function is

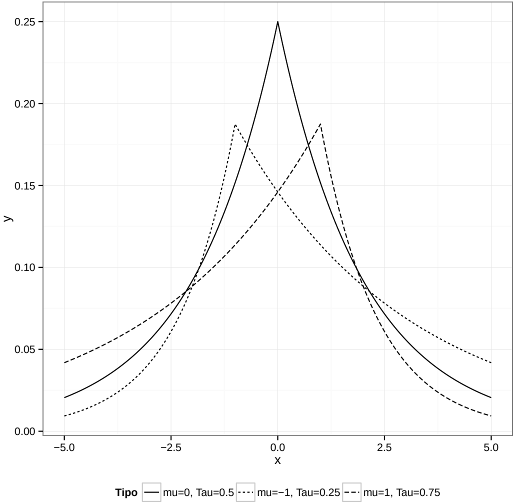
$$f(y; \mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y_i - \mu}{\sigma}\right)\right\}.$$

Location-scale family of distributions:

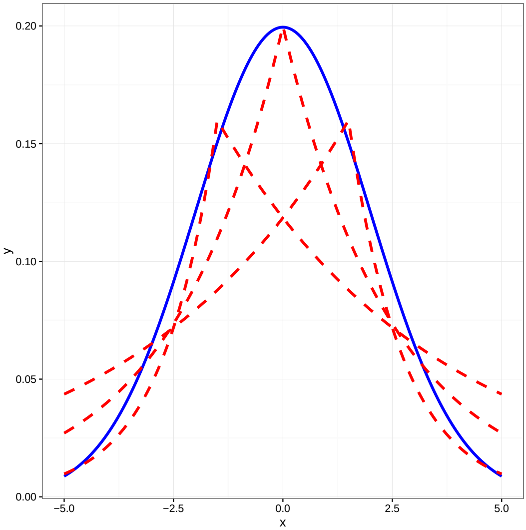
$$X \sim AL(0, 1, \tau) \Rightarrow Y = \mu + \sigma X \sim AL(\mu, \sigma, \tau)$$

- $\mu$ is the $\tau$th quantile of Y

# Densidade da distribuição



# Intuition



# Bayesian quantile regression (First proposal)

ELSEVIER

Statistics & Probability Letters 54 (2001) 437–447

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

## Bayesian quantile regression

Keming Yu, Rana A. Moyeed *

*Department of Mathematics and Statistics, University of Plymouth, Drake circus, Plymouth PL4 8AA, UK*

Received January 2001; received in revised form April 2001

**Abstract**

The paper introduces the idea of Bayesian quantile regression employing a likelihood function that is based on the asymmetric Laplace distribution. It is shown that irrespective of the original distribution of the data, the use of the asymmetric Laplace distribution is a very natural and effective way for modelling Bayesian quantile regression. The paper also demonstrates that improper uniform priors for the unknown model parameters yield a proper joint posterior. The approach is illustrated via a simulated and two real data sets. 

*Keywords:* Asymmetric Laplace distribution; Bayesian inference; Markov chain Monte Carlo methods; Quantile regression

## First proposal

Yu and Moyeed (2001) introduced the Bayesian approach.

$$\pi(\boldsymbol{\beta(\tau)}|\mathbf{y}) \propto L(y|\boldsymbol{\beta(\tau)})\pi(\boldsymbol{\beta(\tau)})$$

- $Y$ is distributed as asymmetric Laplace;
- $\pi(\boldsymbol{\beta(\tau)}) \propto 1$;

Considerations:

- Improper prior distribution $\Rightarrow$ Proper posterior distribution
- Fixed $\sigma = 1$
- MCMC procedure with Metropolis-Hastings

# Bayesian quantile regression (Second proposal)

*Journal of Statistical Computation and Simulation*
Vol. 81, No. 11, November 2011, 1565–1578

Taylor & Francis
Taylor & Francis Group

## Gibbs sampling methods for Bayesian quantile regression

Hideo Kozumi* and Genya Kobayashi

*Graduate School of Business Administration, Kobe University, Rokko, Kobe 657-8501, Japan*

(*Received 27 July 2009; final version received 23 May 2010*)

This paper considers quantile regression models using an asymmetric Laplace distribution from a Bayesian point of view. We develop a simple and efficient Gibbs sampling algorithm for fitting the quantile regression model based on a location-scale mixture representation of the asymmetric Laplace distribution. It is shown that the resulting Gibbs sampler can be accomplished by sampling from either normal or generalized inverse Gaussian distribution. We also discuss some possible extensions of our approach, including the incorporation of a scale parameter, the use of double exponential prior, and a Bayesian analysis of Tobit quantile regression. The proposed methods are illustrated by both simulated and real data.

**Keywords:** asymmetric Laplace distribution; Bayesian quantile regression; double exponential prior; generalized inverse Gaussian distribution; Gibbs sampler; Tobit quantile regression

## Error distributed as asymmetric Laplace

Considering the linear model given by

$$Y_i = \mathbf{x_i}'\boldsymbol{\beta(\tau)} + \epsilon_i$$

where $\epsilon_i$ has asymmetric Laplace distribution.

Kotz et al. (2001):

- This distribution has several mixture representations.

If $\xi$ e $\eta$ are i.i.d. $\mathcal{E}(1)$, then

$$\frac{\xi}{\tau} - \frac{\eta}{1-\tau}$$

has asymmetric Laplace distribution.

# Mixture representation

Kozumi and Kobayashi (2011):

$$Y \sim \mathrm{LA}(\mu, \sigma, \tau)$$

$$\begin{aligned} Y|v &\sim N(\mu + \theta v, \psi^2 \sigma v), \\ v &\sim \mathrm{Exp}(\sigma), \end{aligned}$$

where

$$\theta = \frac{1 - 2\tau}{\tau(1-\tau)}, \quad \psi^2 = \frac{2}{\tau(1-\tau)}.$$

- Allows to obtain more efficient MCMC algorithms.
- One can use a prior normal distribution for $\beta(\tau)$.

## Posterior distributions of $\beta(\tau)$ e $\sigma$

Considering the following prior distributions

$$\begin{aligned} \beta(\tau) &\sim N(b_0, B_0), \\ \sigma &\sim \mathcal{GI}(n_0/2, s_0/2). \end{aligned}$$

Posterior full conditional distributions:

$$\begin{aligned} \beta(\tau)|D, \sigma, v &\sim N(b_1, B_1), \\ \sigma|D, \beta(\tau), v &\sim \mathcal{GI}(n_1/2, s_1/2). \end{aligned}$$

## Posterior distribution of the latent variable

To update the posterior distribution of the latent variables,

$$v_i | D, \beta(\tau), \sigma, v_{(-i)} \propto v_i^{\nu-1} \exp\left\{ -\frac{1}{2} (\delta_i^2 v_i^{-1} + \zeta^2 v_i) \right\}.$$

- Nucleus of the Generalized Inverse Gaussian distribution

$$\nu = \frac{1}{2}, \quad \delta_i^2 = \frac{\big(y_i - x_i'\beta(\tau)\big)^2}{\psi^2 \sigma}, \quad \zeta^2 = \frac{2}{\sigma} + \frac{\theta^2}{\psi^2 \sigma}.$$

- Only $\delta_i^2$ varies with each observation.
- Equals a weighted squared residual.

## Asymmetric Laplace distribution

- Properties of this distribution:

$$E(Y) = \mu + \frac{\sigma(1-2\tau)}{\tau(1-\tau)},$$

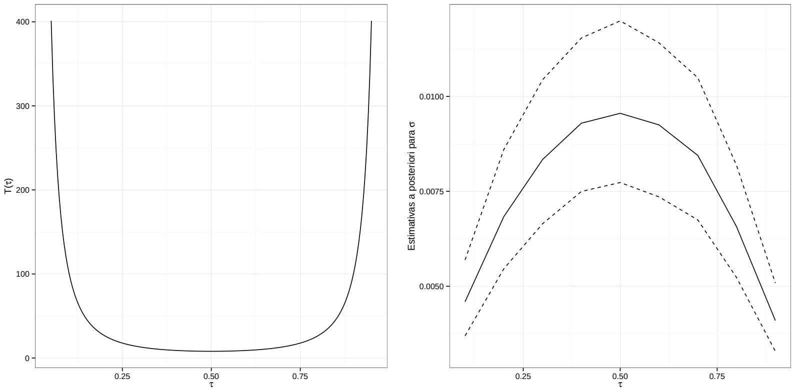
$$\mathrm{Var}(Y) = \sigma^2 T(\tau),$$

where

$$T(\tau) = \frac{(1-2\tau+2\tau^2)}{(1-\tau)^2\tau^2}$$

- Fixing $\sigma = 1$.

# Function $T(\tau)$ and posterior distribution of $\sigma$



# Probability of being an extreme observation

$$O_i = \begin{cases} 1, & \text{if the } i\text{th observation is extreme,} \\ 0, & \text{otherwise.} \end{cases}$$

One could define this probability evaluating using

$$P(O_i = 1) = \frac{1}{n-1} \sum_{j \neq i} P(v_i > v_j | \mathcal{D}).$$

This could be approximated in the MCMC, with

$$P(O_i = 1) = \frac{1}{M} \sum_{l=1}^{M} \mathbb{I}\big(v_i^{(l)} > \max_{k \in 1:M} v_j^{(k)}\big).$$

# Mean posterior probability of being an extreme observation

Observations:

- This probability depends on the quantile.
- There is a greater flexibility in defining extreme observations.
- An observation should be deemed extreme if it is distant from the others.

## Kullback-Leibler Divergence

- Another possibility is to use a divergence measure between the posterior distribution of the latent variables.

Kullback-Leibler divergence:

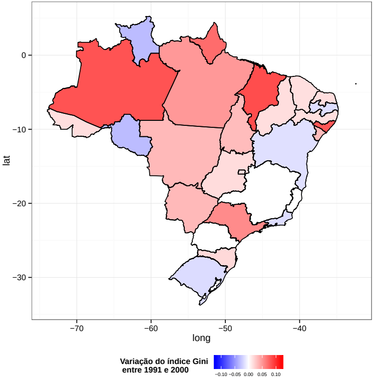
$$K(f_i, f_j) = \int \log\left(\frac{f_i(x)}{f_j(x)}\right) f_i(x)dx.$$

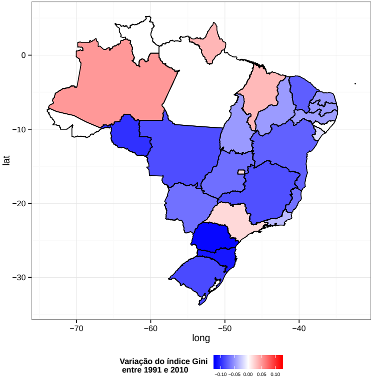
This information could be summarized as a mean value with

$$KL(f_i) = \frac{1}{n-1}\sum_{j\neq i} K(f_i, f_j).$$

# Application

Comparison between Gini indexes of Brazilian states:




Goal: to verify the association between some socio-demographical variables and the Gini index for different quantiles.

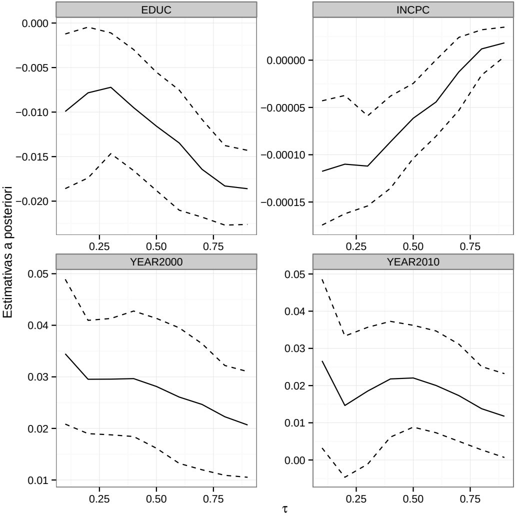
## Proposed model

The following model was proposed to study the conditional quantiles of the Gini index

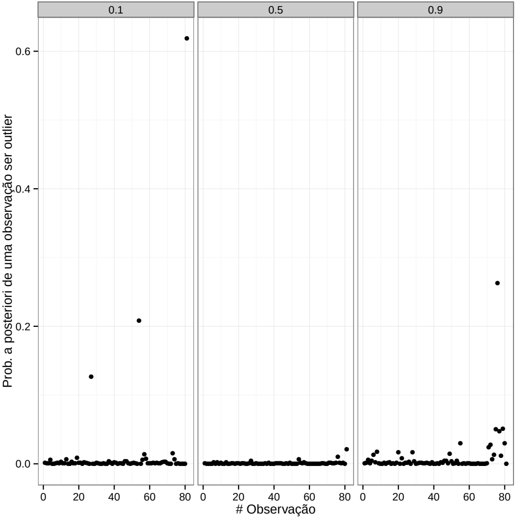
$$Q_{Y_i}(\tau|x_i) = \beta_0(\tau) + \beta_1(\tau)\text{EDUC}_i + \beta_2(\tau)\text{INCPC}_i + \\ \beta_3(\tau)\text{Y2000}_i + \beta_4(\tau)\text{Y2010}_i.$$

- $\beta(\tau) \sim N(0, 100I)$.
- $\sigma \sim IG(3/2, 0.1/2)$.

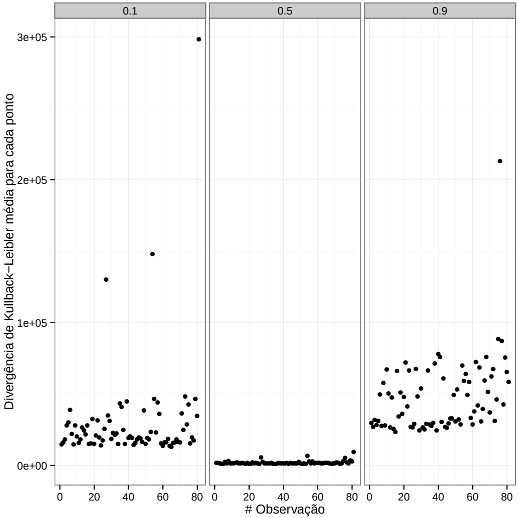
# Coefficient estimates for different quantiles



# Probability of each observation of being extreme



# Kullback-Leibler

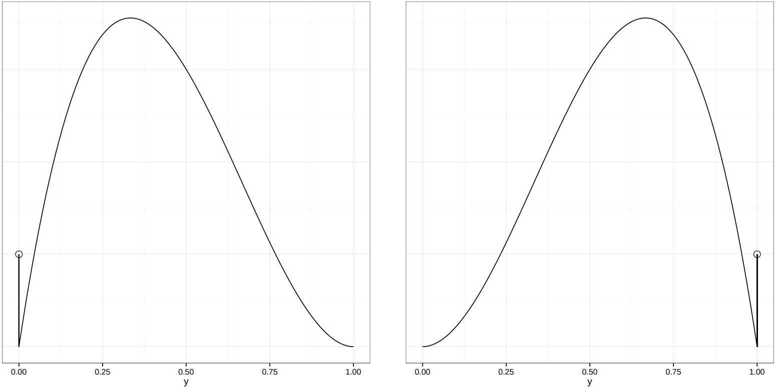


## Consideration on these points

- Three observations from DF with greater values of income per capita.
  - high value of Gini, when these observations were expected to present smaller values.
- One observation of the state of SC, from 2010, that presented the lowest Gini value.
  - this observation presents a great difference for the next value in the sample.
- These observations could be considered extreme in different parts of the conditional distribution of the Gini index.

# Question

How one could model this type of data?




Examples:

- Proportion of households with access to electricity;
- Proportion of votes for a candidate.

## Possibilities

Considering the following density for the response variable:

$$g(y|x,z) = p\mathbb{I}(y = c) + (1-p)f(y|x)\mathbb{I}(0 < y < 1).$$

Ospina and Ferrari (2012):

- beta distribution for the continuous part;
- model $p$ as a function of other variables $z$;
- analyze the mean as a function of other variables $x$.

More complete picture of the conditional distribution:

- Quantile regression for the continuous part.

# Equivariance to monotone transformations property

Let $h(.)$ be a nondecreasing function on $\mathbb{R}$, then for any random variable $Y$,

$$Q_{h(Y)}(\tau|x) = h(Q_Y(\tau|x)),$$

where $Q_Y(\tau|X)$ represents the $\tau$th conditional quantile of $Y$ given $X$.

Then, we could assume that the $\tau$th quantile of transformed response variable follows the linear model

$$Q_{h(Y_i)}(\tau|x_i) = x_i'\beta(\tau),$$

where the link function $h(.)$ maps $[0,1] \to \mathbb{R}$.

## Example with mixture of beta distributions

Considering the following parameterization of the beta distribution,

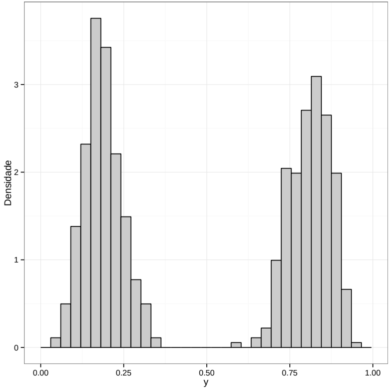
$$f(y_i; \mu_i, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu_i \phi)\Gamma((1-\mu_i)\phi)} y^{\mu_i \phi - 1}(1-y)^{(1-\mu_i)\phi - 1}, \quad 0 < y_i < 1,$$

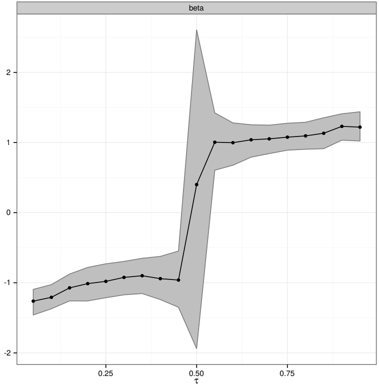
where $E(Y) = \mu$ and $\mathrm{Var}(Y) = (\mu(1-\mu))/(1+\phi)$.

We simulated 600 observations with $\phi = 100$ and from two possible values of $\mu_i$ with equal probability,

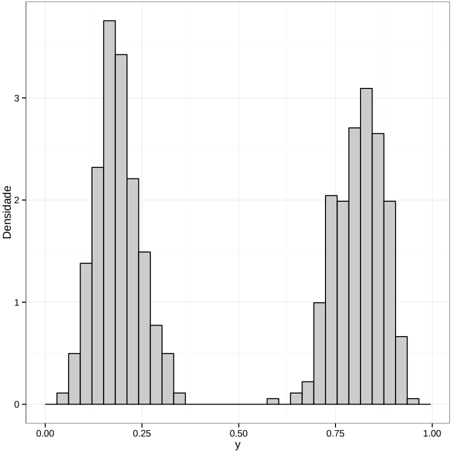
$$\mu_1 = \frac{\exp(-1 - x_i)}{1 + \exp(-1 - x_i)}, \quad \mu_2 = \frac{\exp(1 + x_i)}{1 + \exp(1 + x_i)}.$$

# Resultados


Densidade
y
0.00
0.25
0.50
0.75
1.00
0
1
2
3

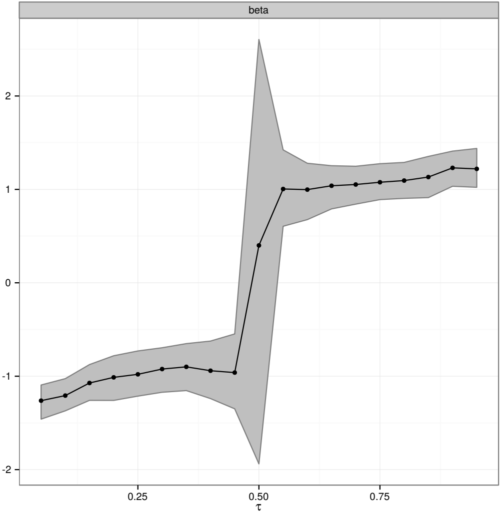

beta
$\tau$
0.25
0.50
0.75
-2
-1
0
1
2

# Histogram of the data




Considering this scheme, we can say $\tau > 0.5$, $\beta(\tau) = 1$ , while for $\tau < 0.5$, $\beta(\tau) = -1$.

# Estimates from posterior distribution considering quantile regression



# Two-part model

Density of the response variable:

$$g(y|x,z) = p\mathbb{I}(y = c) + (1 - p)f(y|x)\mathbb{I}(0 < y < 1).$$

We can model the probability

$$p = P(Y = c|z),$$

using a link function $h_p(.)$, $h_p : (0,1) \to \mathbb{R}$, writing

$$h_p(p) = z'\gamma.$$

## Likelihood and model

Let the sets $C = \{y_i : y_i = c\}$, and $D = \{y_i : 0 < y_i < 1\}$. Then, the augmented likelihood can be written as

$$L(\beta(\tau), \gamma, \sigma, v) = \prod_{y_i \in C} h_p^{-1}(z_i'\gamma) \prod_{y_i \in D} \left(1 - h_p^{-1}(z_i'\gamma)\right) f(y_i),$$

where

$$f(y_i) \propto \left((v_i\sigma)^{-1/2}\right) \exp\left\{-\frac{(y_i - x_i'\beta(\tau) - \theta v_i)^2}{2\psi^2\sigma v_i}\right\}.$$

The full hierarchical model follows from these definitions

$$\begin{aligned}
h(Y_i) &\sim p_i I(Y_i \in C) + (1 - p_i) N(x_i'\beta(\tau) + \theta v_i, \psi^2\sigma v_i) I(Y_i \in D), \\
v_i &\sim \mathcal{E}(\sigma), \\
h_p(p_i) &= z_i'\gamma, \\
\beta(\tau) &\sim N(b_0, B_0), \\
\sigma &\sim \mathcal{IG}(n_0, s_0), \\
\gamma &\sim N(g_0, G_0).
\end{aligned}$$

## Posterior distribution

We are interested in the following posterior distribution,

$$\pi(\beta(\tau), \gamma, \sigma, v|y) \propto L(\beta(\tau), \gamma, \sigma, v)\pi(\beta(\tau))\pi(\gamma)\pi(\sigma)\pi(v).$$

The full conditional distribution for all parameters is

$$\begin{aligned}
\beta(\tau) \mid h(y), v, \gamma, \sigma &\sim N(b_1, B_1), \\
v_i \mid h(y), \beta(\tau), \gamma, \sigma &\sim \mathcal{GIG}(1/2, \hat{\delta}_i, \hat{\xi}_i), \\
\sigma \mid h(y), v, \beta(\tau), \gamma &\sim \mathcal{IG}(\tilde{n}/2, \tilde{s}/2), \\
\pi(\gamma \mid h(y), v, \beta(\tau), \sigma) &\propto \prod_{i \in C} h^{-1}(z_i'\gamma) \prod_{i \in D} \left(1 - h^{-1}(z_i'\gamma)\right) \\
&\quad \times \exp\left\{-\frac{1}{2}(\gamma - g_0)' G_0^{-1}(\gamma - g_0)\right\},
\end{aligned}$$

where $\mathcal{GIG}(\nu, \delta, \zeta)$ represents a generalized inverse Gaussian distribution.

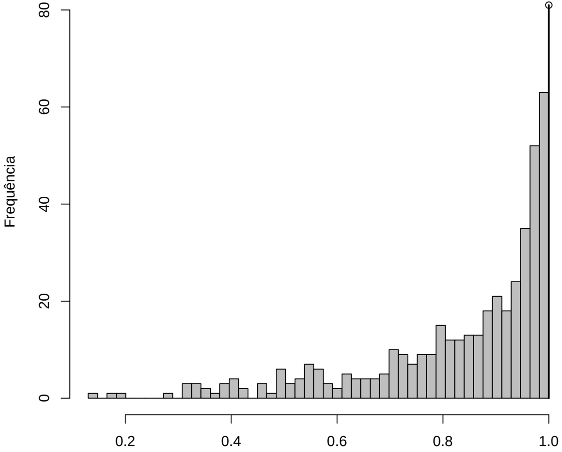
## Analysis of access to electricity in Brazil

- The data consists of 500 cities from the Northeast and Southeast region.
- All variables were measured during a national census in 2000.
- Variables included:
  - proportion of households with access to electricity (PROP_ELEC);
  - region, allowing Southeast to be the reference (REG);
  - population (POP);
  - income per capita (INCPC);
  - human development index (HDI);
  - population density (DENS).

*Model for the continuous part:*

$$h(Q_{y_i}(\tau|x_i)) = \beta_0(\tau) + \beta_1(\tau)\mathrm{REG}_i + \beta_2(\tau)\mathrm{POP}_i + \beta_3(\tau)\mathrm{INCPC}_i + \beta_4(\tau)\mathrm{HDI}_i + \beta_5(\tau)\mathrm{DENS}_i,$$

# Distribution of households with access to electricity in Brazil

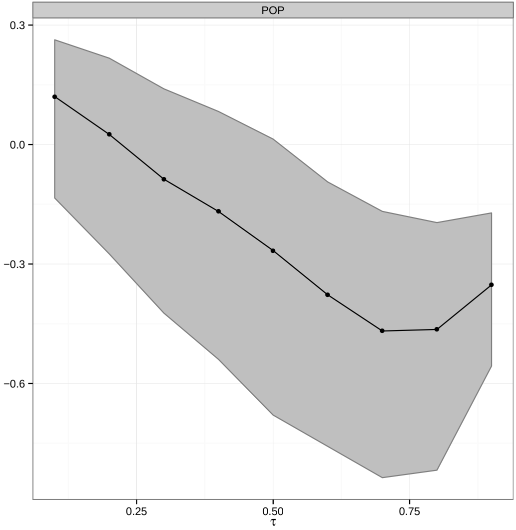


## Model for the probability

$$\log\left(\frac{p_i}{1-p_i}\right) = \gamma_0 + \gamma_1 \mathsf{REG}_i + \gamma_2 \mathsf{POP}_i + \gamma_3 \mathsf{INCPC}_i + \gamma_4 \mathsf{HDI}_i + \gamma_5 \mathsf{DENS}_i,$$

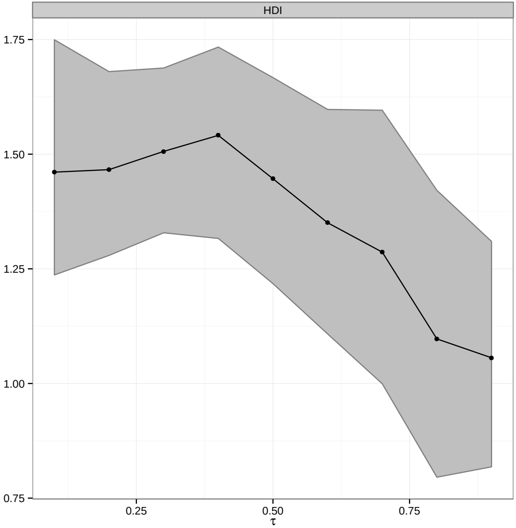
| Covariate | Mean | SD | 95% Credible interval |
|---|---|---|---|
| $\gamma_0$ | -3.94 | 0.68 | [-5.30 ; -2.68] |
| $\gamma_1$ | -1.07 | 1.02 | [-3.12 ; 0.79] |
| $\gamma_2$ | 0.02 | 0.22 | [-0.41 ; 0.48] |
| $\gamma_3$ | 0.71 | 0.42 | [-0.18 ; 1.46] |
| $\gamma_4$ | 2.55 | 0.76 | [1.13 ; 3.88] |
| $\gamma_5$ | 0.85 | 0.35 | [0.22 ; 1.58] |

# Estimates and credibles intervals
## Population



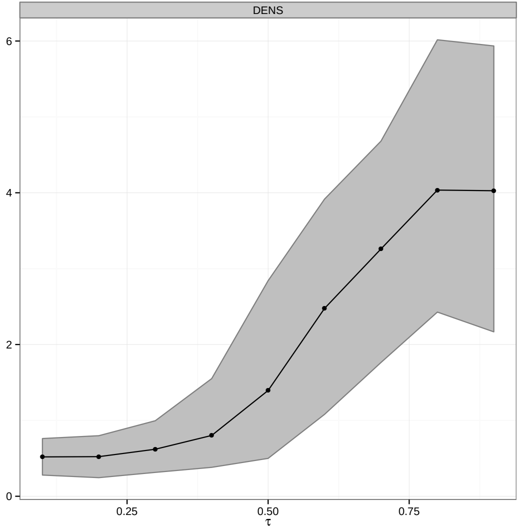
# Estimates and credibles intervals

## Human development index

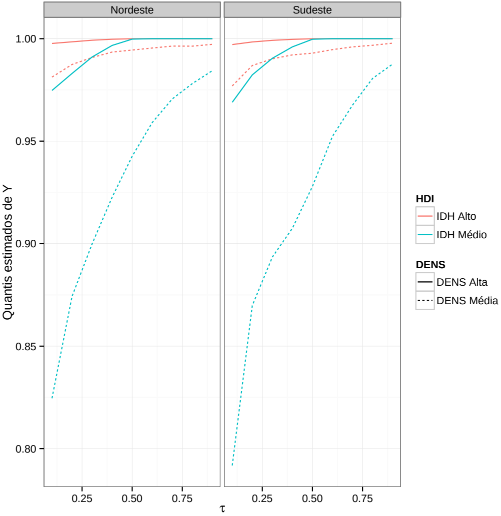


# Estimates and credibles intervals

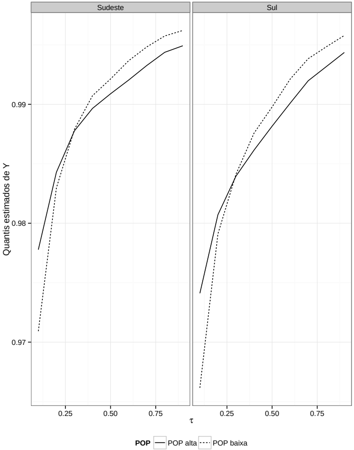
## Population density



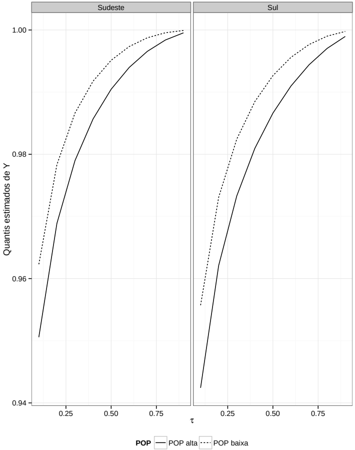
# Estimated quantiles



# Estimated quantiles




**(a)** Quant. Reg



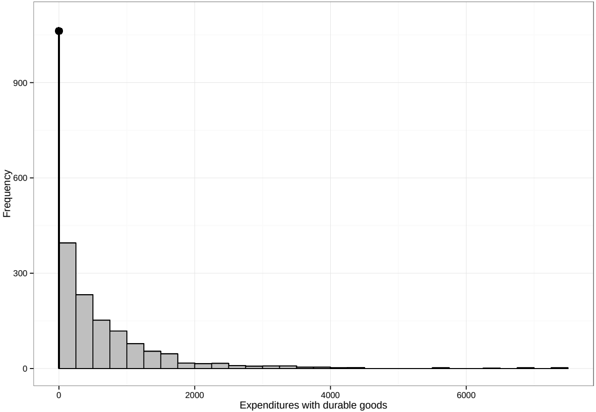

**(b)** Beta Reg.

## Discussion and conclusion

- Quantile regression models are suitable to study proportion data.
- The ALD distribution is a good candidate to explain conditional quantiles for this type of data.
- Our model also deals when data presents a considerable amount of zeros or ones.
- The equivariance property of the quantile function provides a ready framework with this bounded response variable.
- The use of quantile regression models can bring new information about the conditional distribution, other than the conditional mean.

# Continuous data with a discrete component at zero

Expenditures with durable goods in Brazil:



## Some remarks

- Possible models:
  - ◇ Tobit model (Tobin, 1958);
  - ◇ Two-part model (Cragg, 1971).
- What if these two assumptions are valid?
  - ◇ Part of the zeros are censored.
  - ◇ And the other part is a "true zero".
- Quantile regression model:
  - ◇ Conditional quantile as a function of other variables.
  - ◇ Provides more information about the conditional distribution.

## Comparison between the two models

Two-part model

$$g(y) = p\mathbb{I}(y = 0) + (1 - p)f(y)\mathbb{I}(y > 0).$$

Two-part model with censoring at zero

$$g(y) = [p + (1 - p)F(0)]\mathbb{I}(y = 0) + (1 - p)f(y)\mathbb{I}(y > 0)$$

- Part of the observations at zero are censored in the second model
- Asymmetric Laplace distribution is used in the density of the continuous part.

## Censored observations

Kozumi and Kobayashi (2011) considers for censored observations

$$Y^\star \sim NT_{(-\infty,0]}(x'\beta(\tau) + \theta v, \psi^2 \sigma v),$$

If a latent variable $C$ is defined according to

$$C = \begin{cases} 1, & \text{if } Y \text{ is censored,} \\ 0, & \text{otherwise.} \end{cases}$$

The conditional probability of censoring is given by

$$P(C = 1 | Y = 0) = \frac{(1-p)F(0)}{p + (1-p)F(0)},$$

where $F(0)$ can be defined as

$$F(0; \beta(\tau), \sigma) = \begin{cases} \tau \exp\left\{ -\frac{(1-\tau)x'\beta(\tau)}{\sigma} \right\}, & \text{if } x'\beta(\tau) \geqslant 0, \\ 1 - (1-\tau)\exp\left\{ \frac{\tau x'\beta(\tau)}{\sigma} \right\}, & \text{if } x'\beta(\tau) < 0. \end{cases}$$

# Likelihood

Defining the sets

$$C = \{y_i : y_i = 0 \text{ e } c_i = 1\}$$
$$D = \{y_i : y_i = 0 \text{ e } c_i = 0\}$$
$$K = \{y_i : y_i > 0\}$$

And considering the parameters $\xi = (\beta(\tau), \gamma, \sigma)$

$$L(\xi) = \prod_{y_i \in D} \eta^{-1}(z_i'\gamma) \prod_{y_i \in C} (1 - \eta^{-1}(z_i'\gamma)) F(0)$$
$$\prod_{y_i \in K} \left(1 - \eta^{-1}(z_i'\gamma)\right) f(y_i|v_i) f(v_i).$$

## Full conditional distributions

The full conditional distributions for all parameters are

$$
\begin{aligned}
\beta(\tau) \mid y, v, \gamma, \sigma &\sim N(b_1, B_1), \\
v_i \mid h(y), \beta(\tau), \gamma, \sigma &\sim \mathcal{GIG}(1/2, \hat{\delta}_i, \hat{\xi}_i), \\
\sigma \mid y, v, \beta(\tau), \gamma &\sim \mathcal{IG}(\tilde{n}/2, \tilde{s}/2), \\
\pi(\gamma \mid y, v, \beta(\tau), \sigma) &\propto \prod_{i \in D} \eta^{-1}(z_i'\gamma) \prod_{i \in C \cup K} \left(1 - \eta^{-1}(z_i'\gamma)\right) \\
&\times \quad \exp\left\{-\frac{1}{2}(\gamma - g_0)' G_0^{-1}(\gamma - g_0)\right\}.
\end{aligned}
$$

## Latent variable $C$

The posterior update of the latent variable $C$ considers

$$C_i|y, \beta(\tau), \sigma, v \sim Ber\left(\frac{(1-p_i)F(0)}{p_i + (1-p_i)F(0)}\right).$$

- By updating this variable, the MCMC procedure is completed.
- All other parameters are updated according to the two-part model.

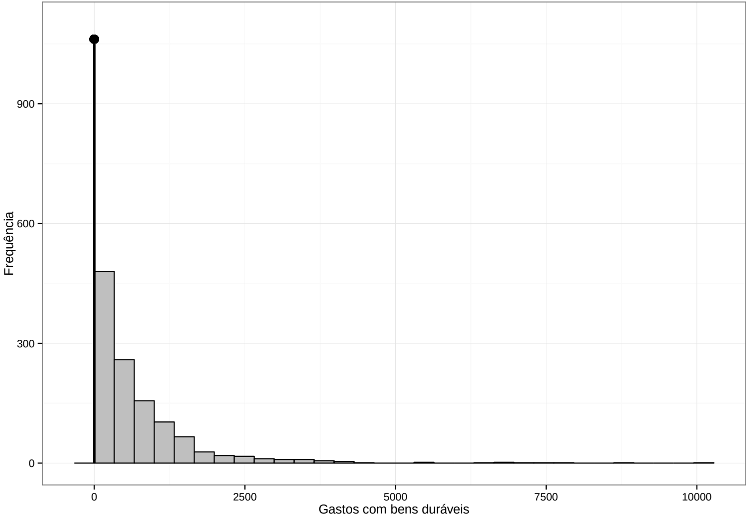
An estimate of the posterior probability of being censored

$$P(C_i = 1|Y, v, \beta(\tau), \sigma, \gamma) = \sum_{k=b+1}^{M} \frac{C_i^{(k)}}{M-b}, \qquad \forall i : y_i \in C \cup D.$$

## Expenditures with durable goods in Brazil

- "Consumer Expenditure Survey" between 2008 and 2009;
- National survey which interviewed around 50.000 households;
- Selecting Maranhão state;
  - 2.240 observations, where 1.062 had no expenditures;
- Used covariates:
  - Gender;
  - Race;
  - Credit card;
  - Age;
  - Years of education;

# Distribution of expenditures with durable goods



## Considered model

- $x_1$: Gender (0: male);
- $x_2$: Race (0: white);
- $x_3$: Age;
- $x_4$: Years of education;
- $x_5$: Credit card (0: has);

$$\log\left(\frac{p_i}{1-p_i}\right) = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i3} + \gamma_4 x_{i4} + \gamma_5 x_{i5},$$

$$Q_{\sqrt{Y_i}}(\tau|x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}.$$

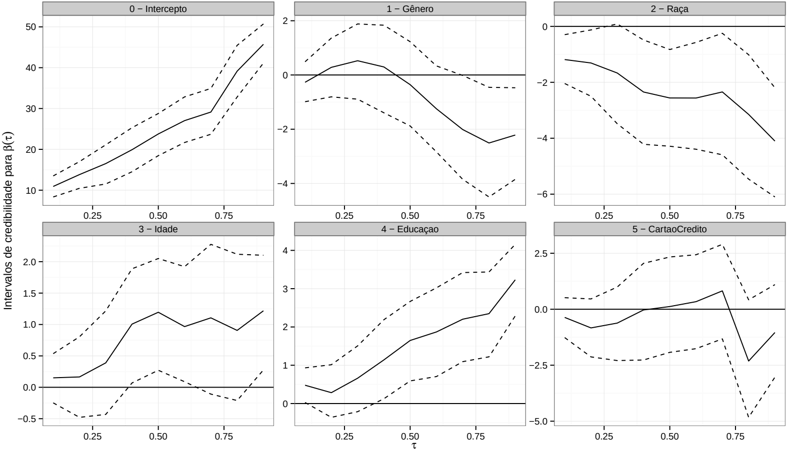
# Coefficient estimates for the probability

Estimates for the model when $\tau = 0,50$:

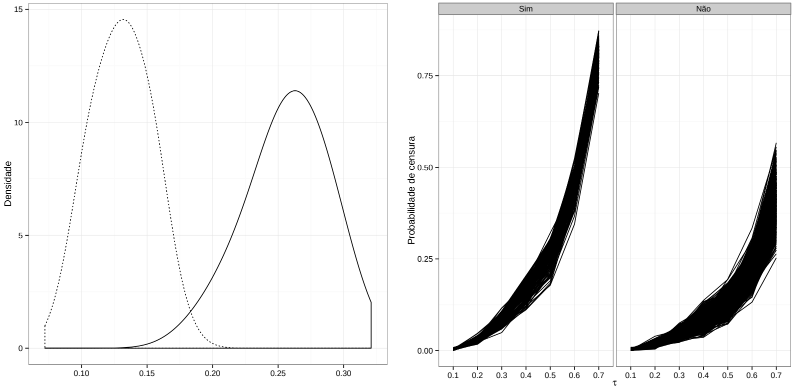
| Variabel | Estimate | Credible interval |
|---|---|---|
| Intercept | -2,14 | [-2,88 ; -1,49] |
| Gender | 0,02 | [-0,16 ; 0,18] |
| Race | 0,16 | [-0,05 ; 0,36] |
| Age | 0,08 | [-0,02 ; 0,18] |
| Education | -0,12 | [-0,23 ; -0,01] |
| Credit card | 0,79 | [0,50 ; 1,09] |

- Education has a negative effect on the odds of zero expenditures.
- People without credit card have 2 times the odds of not having any expenditures in comparison with people with credit card.

# Point estimates and posterior intervals



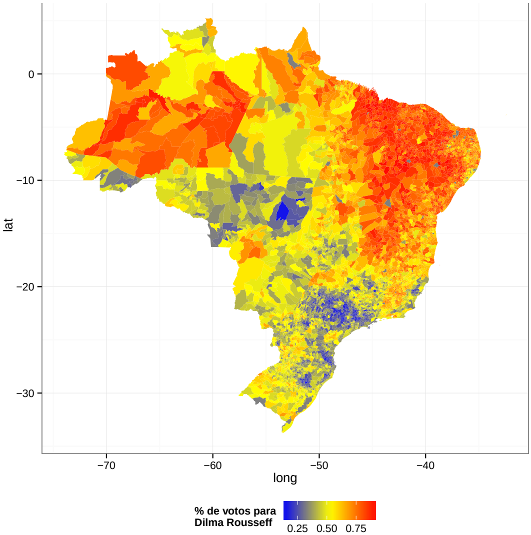
# Probabilities of being censored given credit card




(a) Probability of being censored in the median model. Yes = solid line; no = dashed line.

(b) Profile of these probabilities for several quantiles.

# Analysis of the presidential elections in Brazil via Bayesian spatial quantile regression



# Motivation

- Election with tight result:
  - ◇ Which variables could be related to the distribution of votes?
- Response variable presented as proportion data:
  - ◇ Beta regression:
    - Conditional mean as a function of other covariates.
  - ◇ Quantile regression:
    - Conditional quantile as a function of other covariates.
- Spatial representation of data:
  - ◇ Asymetric Laplace process

## Bayesian spatial quantile regression model

Lum and Gelfand (2012) define Asymetric Laplace process (ALP)

$$
\begin{aligned}
Y(s) &= X(s)\beta(\tau) + \epsilon(s), \\
\epsilon(s) &= \theta v(s) + \sqrt{\psi^2 \sigma v(s)} Z(s), \\
Z(s) &\sim GP(0, C(s; \lambda)), \\
v(s) &\sim \mathrm{Exp}(\sigma),
\end{aligned}
$$

In the covariance function, one can use the squared exponential function

$$c(s, s'; \lambda) = \exp\{-\lambda ||s - s'||^2\}$$

# Model fitting

The full conditional posterior distribution of $v_i$ is proportional to

$$v_i^{-1/2} \exp\left\{ -\frac{1}{2\sigma\psi^2} u^t D\left(\sqrt{v}\right) C^{-1} D\left(\sqrt{v}\right) u - \frac{v_i}{\sigma} \right\},$$

where $u$ is a vector $n \times 1$, where each term is equal to

$$u_i = Y_i - x_i' \beta(\tau) - \theta v_i.$$

Lum and Gelfand (2012) suggest to add a parameter $\alpha \in (0, 1)$ in the following way

$$c(s_i, s_j; \lambda, \alpha) = (1 - \alpha) \exp\{-\lambda ||s_i - s_j||^2\} + \alpha \mathbb{I}(i = j),$$

where $\lambda > 0$ e $\alpha \in (0, 1)$

# Asymetric Laplace predictive process (ALPP)

To define the ALPP, is necessary to redefine $\epsilon(s)$ as

$$
\begin{aligned}
\ddot{\epsilon}(s) &= \theta v(s) + \sqrt{\psi^2 \sigma v(s)}\ddot{Z}(s) \\
\ddot{Z}(s) &= \ddot{w}(s) + \delta(s) \\
\ddot{w}(s) &= \tilde{w}(s) + \tilde{\delta}(s) \\
\tilde{w}(s) &= E(Z(s)|Z^*) = c^t(s, s^*; \lambda){C^*}^{-1} Z^* \\
Z^* &\sim N_m(0, (1-\alpha)C(s^*, {s^*}'; \lambda)) \\
\tilde{\delta}(s) &\sim N(0, \ddot{\sigma}(s)) \\
\ddot{\delta}(s) &= (1-\alpha)(c(s, s; \lambda) - c^t(s, s^*; \lambda){C^*}^{-1} c(s, s^*; \lambda)) \\
\delta(s) &\sim N(0, \alpha),
\end{aligned}
$$

## Asymetric Laplace predictive process (ALPP)

To define the ALPP, is necessary to redefine $\epsilon(s)$ as

$$
\begin{aligned}
\ddot{\epsilon}(s) &= \theta v(s) + \sqrt{\psi^2 \sigma v(s)}\ddot{Z}(s) \\
\ddot{Z}(s) &= \ddot{w}(s) + \delta(s) \\
\ddot{w}(s) &= \tilde{w}(s) + \tilde{\delta}(s) \\
\tilde{w}(s) &= E(Z(s)|Z^*) = c^t(s, s^*; \lambda) {C^*}^{-1} Z^* \\
Z^* &\sim N_m(0, (1-\alpha) C(s^*, {s^*}'; \lambda)) \\
\tilde{\delta}(s) &\sim N(0, \ddot{\sigma}(s)) \\
\ddot{\sigma}(s) &= (1-\alpha)(c(s, s; \lambda) - c^t(s, s^*; \lambda) {C^*}^{-1} c(s, s^*; \lambda)) \\
\delta(s) &\sim N(0, \alpha),
\end{aligned}
$$

## Asymetric Laplace predictive process

If one uses that

$$\ddot{\epsilon}(s) \sim N(\theta v(s), \psi^2 \sigma D\left(\sqrt{v}\right) \Sigma_\xi D\left(\sqrt{v}\right)),$$

where

$$\Sigma_\xi = (1-\alpha) c^t(s, s^*; \lambda) {C^*}^{-1} c(s, s^*; \lambda) + D(\alpha + \ddot{\sigma}(s)),$$

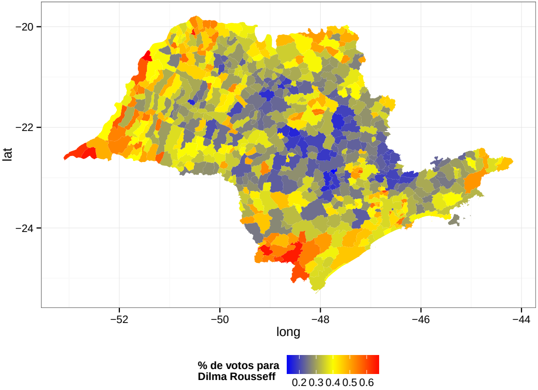
and $D(\alpha + \ddot{\sigma}(s))$ is a diagonal matrix.

To invert matrix $\Sigma_\xi$, is possible to use Sherman-Morrison-Woodbury formula which states that

$$(A + UWV)^{-1} = A^{-1} - A^{-1} U (W^{-1} + VA^{-1} U)^{-1} VA^{-1}.$$

# Presidential election in Brazil

Utilizing only the votes of the state of São Paulo.



## Proposed model

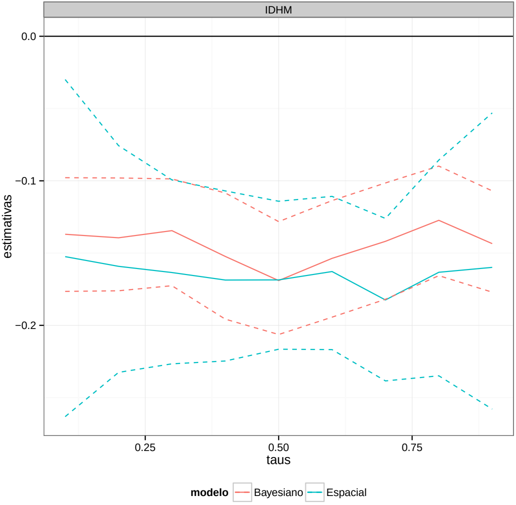
The following model was proposed to study the proportion of votes

$$Q_{h(y)}(\tau|x) = \beta_0(\tau)+\beta_1(\tau)x_1+\beta_2(\tau)x_2+\beta_3(\tau)x_3+\beta_4(\tau)x_4+\beta_5(\tau)x_5,$$

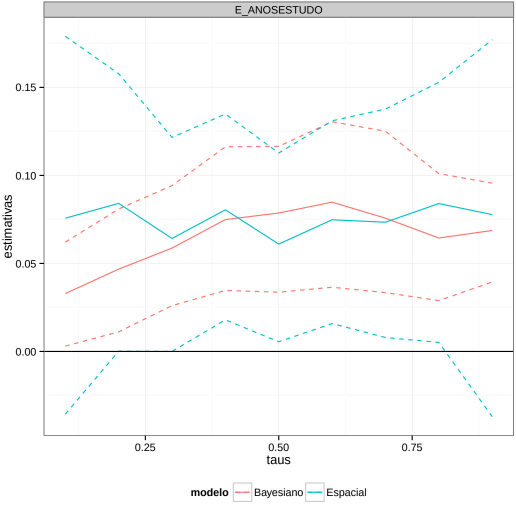
where

- $x_1$ human development index (IDHM),
- $x_2$: average of years of education (E_ANOSESTUDO),
- $x_3$ variation of Gini between years 2000 and 2010 (GINI_DELTA),
- $x_4$ variation of income per capita years 2000 and 2010 (RDPC_DELTA),
- $x_5$ average value received by family in Cash Transfer program, Bolsa Família.

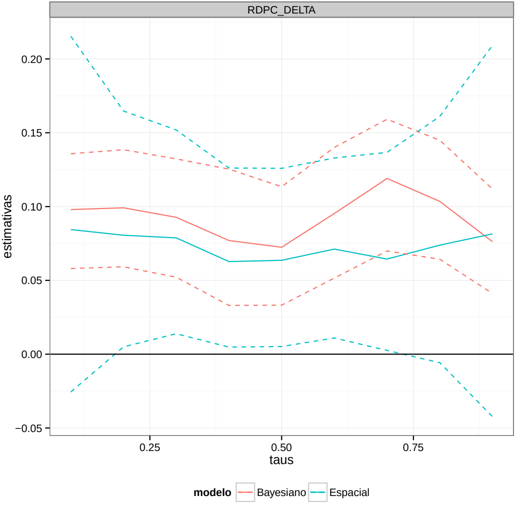
# Coefficient estimates



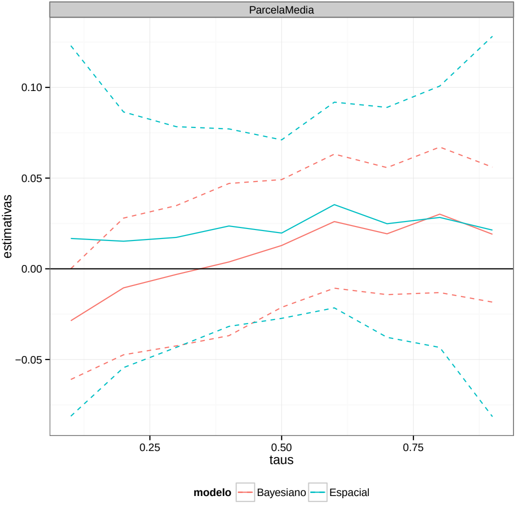
# Coefficient estimates



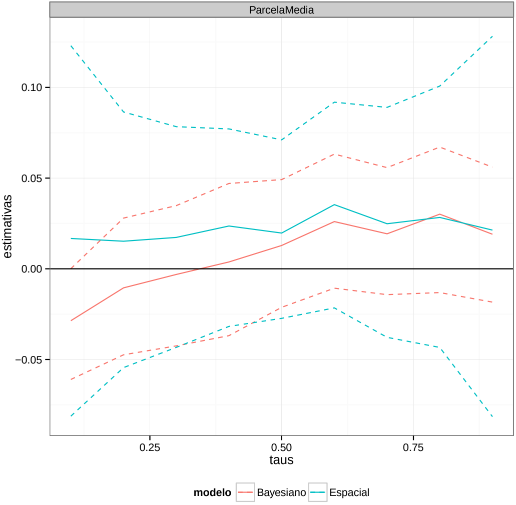
# Coefficient estimates



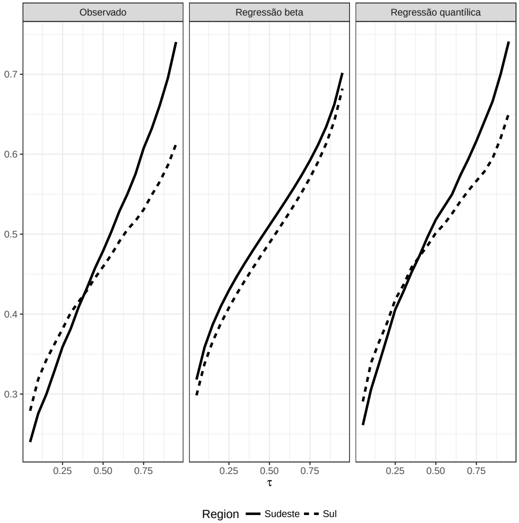
# Coefficient estimates



# Coefficient estimates



# Comparison of fit models



# R package - baquantreg

**R package:** baquantreg.
Available on Github.

# Comentários

- Quantile regression models can be used for proportion data.
- Equivariance property allows the use of the asymmetric Laplace.
- The two-part model can be extended to the presence of zeros and ones simultaneously.
- Censoring probability depends of the quantile of interest.
- Extension of the Tobit quantile regression model with censoring at zero.

# ¡Gracias!

brunorsantos@ufba.br