

# Quantile and Expectile Regressions

Arthur Charpentier (Université de Rennes 1)

Erasmus, May 2017

<http://freakonometrics.hypotheses.org>



## Econometrics vs Machine Learning

proba loss

finir par : but sometimes, loss functions yield a probabilistic interpretation

## Elicitable Measures

“**elicitable**” means “being a minimizer of a suitable expected score”

$T$  is an elicitable function if there exists a scoring function  $S : \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty)$  such that

$$T(Y) = \operatorname{argmin}_{x \in \mathbb{R}} \left\{ \int_{\mathbb{R}} S(x, y) dF(y) \right\} = \operatorname{argmin}_{x \in \mathbb{R}} \left\{ \mathbb{E}[S(x, Y)] \text{ where } Y \sim F. \right\}$$

see Gneiting (2011) [Making and evaluating point forecasts](#).

Example: **mean**,  $T(Y) = \mathbb{E}[Y]$  is elicited by  $S(x, y) = \|x - y\|_{\ell_2}^2$

Example: **median**,  $T(Y) = \text{median}[Y]$  is elicited by  $S(x, y) = \|x - y\|_{\ell_1}$

Example: **quantile**,  $T(Y) = Q_Y(\tau)$  is elicited by

$$S(x, y) = \tau(y - x)_+ + (1 - \tau)(y - x)_-$$

Example: **expectile**,  $T(Y) = E_Y(\tau)$  is elicited by

$$S(x, y) = \tau(y - x)_+^2 + (1 - \tau)(y - x)_-^2$$

## OLS Regression, $\ell_2$ norm and Expected Value

Let  $\mathbf{y} \in \mathbb{R}^d$ ,  $\bar{y} = \operatorname{argmin}_{m \in \mathbb{R}} \left\{ \sum_{i=1}^n \frac{1}{n} \underbrace{[y_i - m]}_{\varepsilon_i}^2 \right\}$ . It is the empirical version of

$$\mathbb{E}[Y] = \operatorname{argmin}_{m \in \mathbb{R}} \left\{ \int \underbrace{[y - m]}_{\varepsilon}^2 dF(y) \right\} = \operatorname{argmin}_{m \in \mathbb{R}} \left\{ \mathbb{E}[\|Y - m\|_{\ell_2}] \right\}$$

where  $Y$  is a random variable.

Thus,  $\operatorname{argmin}_{m(\cdot): \mathbb{R}^k \rightarrow \mathbb{R}} \left\{ \sum_{i=1}^n \frac{1}{n} \underbrace{[y_i - m(\mathbf{x}_i)]}_{\varepsilon_i}^2 \right\}$  is the empirical version of  $\mathbb{E}[Y | \mathbf{X} = \mathbf{x}]$ .

See Legendre (1805) *Nouvelles méthodes pour la détermination des orbites des comètes* and Gauß (1809) *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*.

## Median Regression, $\ell_1$ norm and Median

Let  $\mathbf{y} \in \mathbb{R}^d$ ,  $\text{median}[\mathbf{y}] \in \operatorname{argmin}_{m \in \mathbb{R}} \left\{ \sum_{i=1}^n \frac{1}{n} \underbrace{|y_i - m|}_{\varepsilon_i} \right\}$ . It is the empirical version of

$$\text{median}[Y] \in \operatorname{argmin}_{m \in \mathbb{R}} \left\{ \int \underbrace{|y - m|}_{\varepsilon} dF(y) \right\} = \operatorname{argmin}_{m \in \mathbb{R}} \left\{ \mathbb{E} \left[ \underbrace{\|Y - m\|_{\ell_1}}_{\varepsilon} \right] \right\}$$

where  $Y$  is a random variable,  $\mathbb{P}[Y \leq \text{median}[Y]] \geq \frac{1}{2}$  and  $\mathbb{P}[Y \geq \text{median}[Y]] \geq \frac{1}{2}$ .

$\operatorname{argmin}_{m(\cdot): \mathbb{R}^k \rightarrow \mathbb{R}} \left\{ \sum_{i=1}^n \frac{1}{n} \underbrace{|y_i - m(\mathbf{x}_i)|}_{\varepsilon_i} \right\}$  is the empirical version of  $\text{median}[Y | \mathbf{X} = \mathbf{x}]$ .

See Boscovich (1757) *De Litteraria expeditione per pontificiam ditionem ad dimetiendos duos meridiani* and Laplace (1793) *Sur quelques points du système du monde*.

## Quantiles and Expectiles

Consider the following risk functions

$$\mathcal{R}_\tau^q(u) = u \cdot (\tau - \mathbf{1}(u < 0)), \quad \tau \in [0, 1]$$

with  $\mathcal{R}_{1/2}^q(u) \propto |u| = \|u\|_{\ell_1}$ , and

$$\mathcal{R}_\tau^e(u) = u^2 \cdot (\tau - \mathbf{1}(u < 0)), \quad \tau \in [0, 1]$$

with  $\mathcal{R}_{1/2}^e(u) \propto u^2 = \|u\|_{\ell_2}^2$ .

$$Q_Y(\tau) = \operatorname{argmin}_m \left\{ \mathbb{E}(\mathcal{R}_\tau^q(Y - m)) \right\}$$

which is the median when  $\tau = 1/2$ ,

$$E_Y(\tau) = \operatorname{argmin}_m \left\{ \mathbb{E}(\mathcal{R}_\tau^e(X - m)) \right\}$$

which is the expected value when  $\tau = 1/2$ .

## Quantiles and Expectiles

One can also write

$$\text{quantile: } \operatorname{argmin} \left\{ \sum_{i=1}^n \omega_{\tau}^q(\varepsilon_i) \left| \underbrace{y_i - q_i}_{\varepsilon_i} \right| \right\} \text{ where } \omega_{\tau}^q(\epsilon) = \begin{cases} 1 - \tau & \text{if } \epsilon \leq 0 \\ \tau & \text{if } \epsilon > 0 \end{cases}$$

$$\text{expectile: } \operatorname{argmin} \left\{ \sum_{i=1}^n \omega_{\tau}^e(\varepsilon_i) \left( \underbrace{y_i - q_i}_{\varepsilon_i} \right)^2 \right\} \text{ where } \omega_{\tau}^e(\epsilon) = \begin{cases} 1 - \tau & \text{if } \epsilon \leq 0 \\ \tau & \text{if } \epsilon > 0 \end{cases}$$

Expectiles are unique, not quantiles...

Quantiles satisfy  $\mathbb{E}[\operatorname{sign}(Y - Q_Y(\tau))] = 0$

Expectiles satisfy  $\tau \mathbb{E}[(Y - e_Y(\tau))_+] = (1 - \tau) \mathbb{E}[(Y - e_Y(\tau))_-]$

(those are actually the first order conditions of the optimization problem).

## Quantiles and $M$ -Estimators

There are connections with  $M$ -estimators, as introduced in Serfling (1980) *Approximation Theorems of Mathematical Statistics*, chapter 7.

For any function  $h(\cdot, \cdot)$ , the  $M$ -functional is the solution  $\beta$  of

$$\int h(y, \beta) dF_Y(y) = 0$$

, and the  $M$ -estimator is the solution of

$$\int h(y, \beta) d\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^n h(y_i, \beta) = 0$$

Hence, if  $h(y, \beta) = y - \beta$ ,  $\beta = \mathbb{E}[Y]$  and  $\hat{\beta} = \bar{y}$ .

And if  $h(y, \beta) = \mathbf{1}(y < \beta) - \tau$ , with  $\tau \in (0, 1)$ , then  $\beta = F_Y^{-1}(\tau)$ .

## Quantiles, Maximal Correlation and Hardy-Littlewood-Polya

If  $x_1 \leq \dots \leq x_n$  and  $y_1 \leq \dots \leq y_n$ , then  $\sum_{i=1}^n x_i y_i \geq \sum_{i=1}^n x_i y_{\sigma(i)}$ ,  $\forall \sigma \in \mathcal{S}_n$ , and  $\mathbf{x}$  and  $\mathbf{y}$  are said to be comonotonic.

The continuous version is that  $X$  and  $Y$  are comonotonic if

$$\mathbb{E}[XY] \geq \mathbb{E}[X\tilde{Y}] \text{ where } \tilde{Y} \stackrel{\mathcal{L}}{=} Y,$$

One can prove that

$$Y = Q_Y(F_X(X)) = \operatorname{argmax}_{\tilde{Y} \sim F_Y} \{\mathbb{E}[X\tilde{Y}]\}$$

## Expectiles as Quantiles

For every  $Y \in L^1$ ,  $\tau \mapsto e_Y(\tau)$  is continuous, and strictly increasing

if  $Y$  is absolutely continuous,  $\frac{\partial e_Y(\tau)}{\partial \tau} = \frac{\mathbb{E}[|X - e_Y(\tau)|]}{(1 - \tau)F_Y(e_Y(\tau)) + \tau(1 - F_Y(e_Y(\tau)))}$

if  $X \leq Y$ , then  $e_X(\tau) \leq e_Y(\tau) \forall \tau \in (0, 1)$

“Expectiles have properties that are similar to quantiles” Newey & Powell (1987)

**Asymmetric Least Squares Estimation and Testing.** The reason is that expectiles of a distribution  $F$  are quantiles of a distribution  $G$  which is related to  $F$ , see Jones (1994) **Expectiles and M-quantiles are quantiles:** let

$$G(t) = \frac{P(t) - tF(t)}{2[P(t) - tF(t)] + t - \mu} \text{ where } P(s) = \int_{-\infty}^s ydF(y).$$

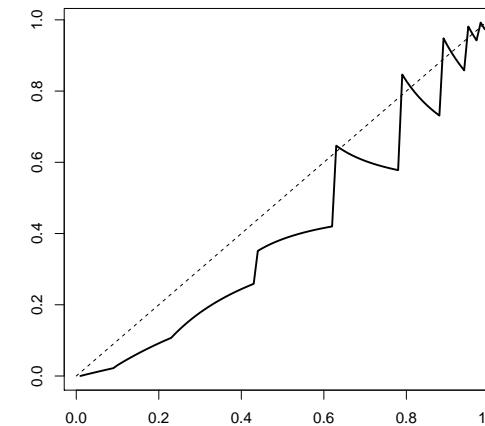
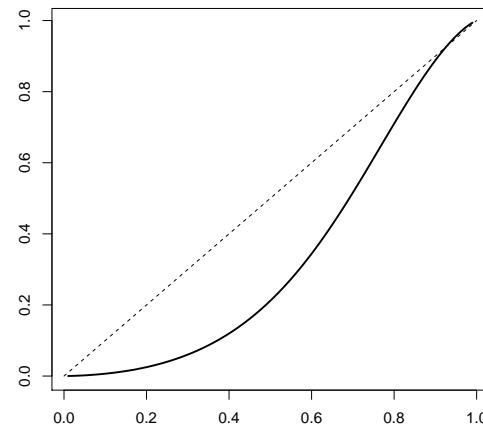
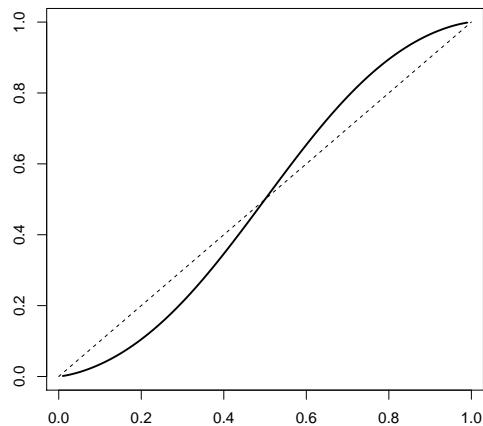
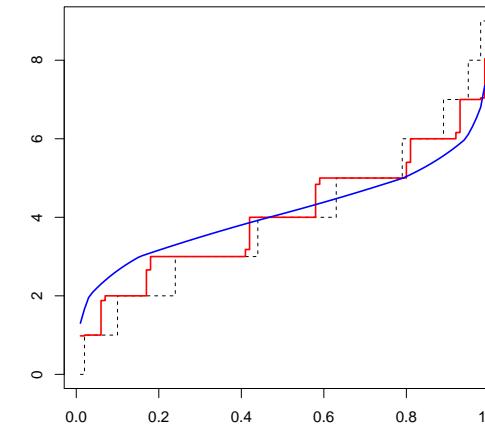
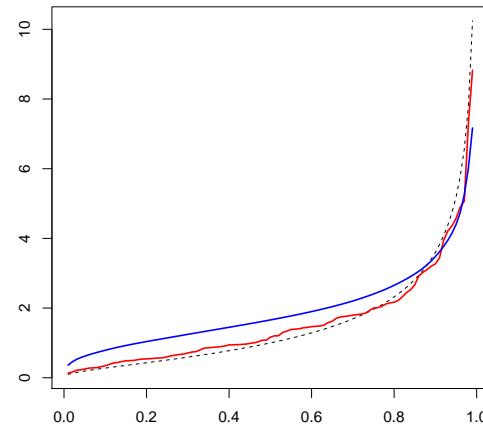
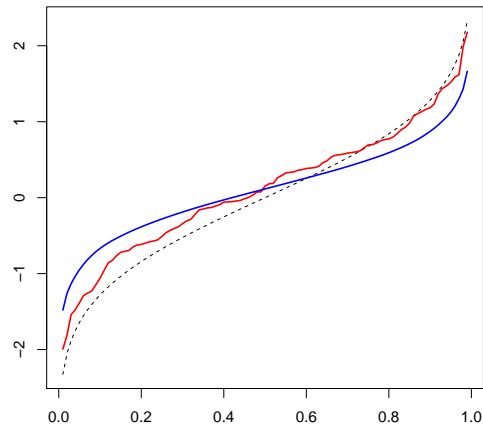
The expectiles of  $F$  are the quantiles of  $G$ .

```

1 > x <- rnorm(99)
2 > library(expectreg)
3 > e <- expectile(x, probs = seq(0, 1, 0.1))

```

## Expectiles as Quantiles



## Empirical Quantiles

Consider some i.id. sample  $\{y_1, \dots, y_n\}$  with distribution  $F$ . Set

$$Q_\tau = \operatorname{argmin} \left\{ \mathbb{E}[\mathcal{R}_\tau^q(Y - q)] \right\} \text{ where } Y \sim F \text{ and } \hat{Q}_\tau \in \operatorname{argmin} \left\{ \sum_{i=1}^n \mathcal{R}_\tau^q(y_i - q) \right\}$$

Then as  $n \rightarrow \infty$

$$\sqrt{n}(\hat{Q}_\tau - Q_\tau) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\tau(1-\tau)}{f^2(Q_\tau)}\right)$$

Sketch of the proof:  $y_i = Q_\tau + \varepsilon_i$ , set  $h_n(q) = \frac{1}{n} \sum_{i=1}^n (\mathbf{1}(y_i < q) - \tau)$ , which is a non-decreasing function, with

$$\mathbb{E} \left[ Q_\tau + \frac{u}{\sqrt{n}} \right] = F_Y \left( Q_\tau + \frac{u}{\sqrt{n}} \right) \sim f_Y(Q_\tau) \frac{u}{\sqrt{n}}$$

$$\operatorname{Var} \left[ Q_\tau + \frac{u}{\sqrt{n}} \right] \sim \frac{F_Y(Q_\tau)[1 - F_Y(Q_\tau)]}{n} = \frac{\tau(1-\tau)}{n}.$$

## Empirical Expectiles

Consider some i.id. sample  $\{y_1, \dots, y_n\}$  with distribution  $F$ . Set

$$\mu_\tau = \operatorname{argmin} \left\{ \mathbb{E}[\mathcal{R}_\tau^e(Y - m)] \right\} \text{ where } Y \sim F \text{ and } \hat{\mu}_\tau = \operatorname{argmin} \left\{ \sum_{i=1}^n \mathcal{R}_\tau^e(y_i - m) \right\}$$

Then as  $n \rightarrow \infty$

$$\sqrt{n}(\hat{\mu}_\tau - \mu_\tau) \xrightarrow{\mathcal{L}} \mathcal{N}(0, s^2)$$

for some  $s^2$ , if  $\operatorname{Var}[Y] < \infty$ . Define the identification function

$$\mathcal{I}_\tau(x, y) = \tau(y - x)_+ + (1 - \tau)(y - x)_- \quad (\text{elicitable score for quantiles})$$

so that  $\mu_\tau$  is solution of  $\mathbb{E}[\mathcal{I}(\mu_\tau, Y)] = 0$ . Then

$$s^2 = \frac{\mathbb{E}[\mathcal{I}(\mu_\tau, Y)^2]}{(\tau[1 - F(\mu_\tau)] + [1 - \tau]F(\mu_\tau))^2}.$$

## Quantile Regression

We want to solve, here,  $\min \left\{ \sum_{i=1}^n \mathcal{R}_\tau^q(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \right\}$

$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$  so that  $\hat{Q}_{y|\mathbf{x}}(\tau) = \mathbf{x}^\top \hat{\boldsymbol{\beta}} + F_\varepsilon^{-1}(\tau)$

## Geometric Properties of the Quantile Regression

Observe that the median regression will always have two supporting observations.

Start with some regression line,  $y_i = \beta_0 + \beta_1 x_i$

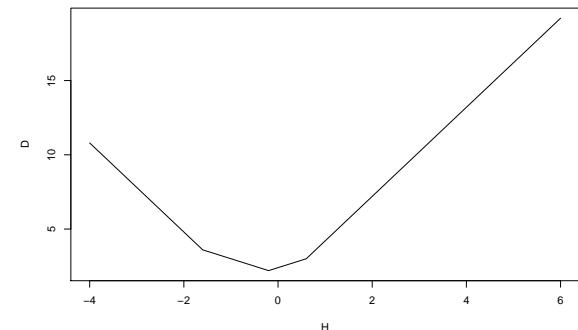
Consider small translations  $y_i = (\beta_0 \pm \epsilon) + \beta_1 x_i$

We minimize  $\sum_{i=1}^n |y_i - (\beta_0 + \beta_1 x_i)|$

From line blue, a shift up decrease the sum by  $\epsilon$   
until we meet point on the left

an additional shift up will increase the sum

We will necessarily pass through one point  
(observe that the sum is piecewise linear in  $\epsilon$ )



## Geometric Properties of the Quantile Regression

Consider now rotations of the line around the support point

If we rotate up, we increase the sum of absolute difference (large impact on the point on the right)

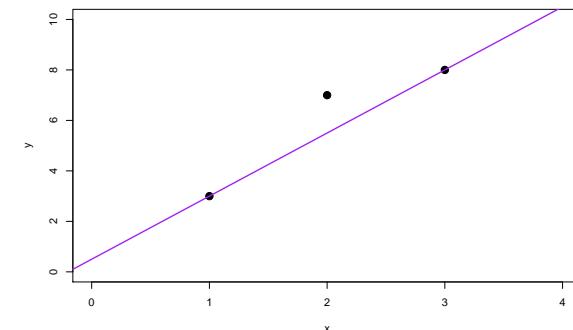
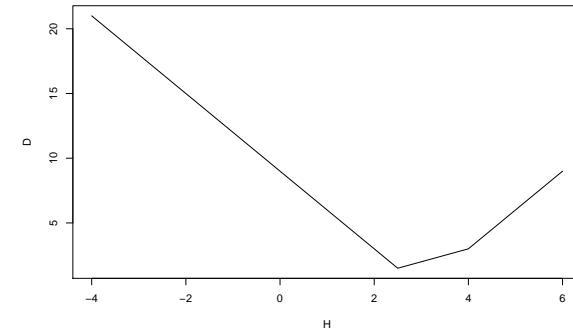
If we rotate down, we decrease the sum, until we reach the point on the right

Thus, the median regression will always have two supporting observations.

```

1 > library(quantreg)
2 > fit <- rq(dist ~ speed, data=cars, tau=.5)
3 > which(predict(fit) == cars$dist)
4 1 21 46
5 1 21 46

```



## Distributional Aspects

OLS are equivalent to MLE when  $Y - m(\mathbf{x}) \sim \mathcal{N}(0, \sigma^2)$ , with density

$$g(\epsilon) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right)$$

Quantile regression is equivalent to Maximum Likelihood Estimation when  $Y - m(\mathbf{x})$  has an asymmetric Laplace distribution

$$g(\epsilon) = \frac{\sqrt{2}}{\sigma} \frac{\kappa}{1 + \kappa^2} \exp\left(-\frac{\sqrt{2}\kappa \mathbf{1}(\epsilon > 0)}{\sigma\kappa \mathbf{1}(\epsilon < 0)} |\epsilon|\right)$$

## Quantile Regression and Iterative Least Squares

start with some  $\beta^{(0)}$  e.g.  $\beta^{\text{ols}}$

at stage  $k$  :

$$\text{let } \varepsilon_i^{(k)} = y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(k-1)}$$

$$\text{define weights } \omega_i^{(k)} = \mathcal{R}_\tau(\varepsilon_i^{(k)})$$

compute weighted least square to estimate  $\beta^{(k)}$

One can also consider a smooth approximation of  $\mathcal{R}_\tau^q(\cdot)$ , and then use Newton-Raphson.

## Optimization Algorithm

Primal problem is

$$\min_{\beta, \mathbf{u}, \mathbf{v}} \left\{ \tau \mathbf{1}^\top \mathbf{u} + (1 - \tau) \mathbf{1}^\top \mathbf{v} \right\} \text{ s.t. } \mathbf{y} = \mathbf{X}\beta + \mathbf{u} - \mathbf{v}, \text{ with } \mathbf{u}, \mathbf{v} \in \mathbb{R}_+^n$$

and the dual version is

$$\max_{\mathbf{d}} \left\{ \mathbf{y}^\top \mathbf{d} \right\} \text{ s.t. } \mathbf{X}^\top \mathbf{d} = (1 - \tau) \mathbf{X}^\top \mathbf{1} \text{ with } \mathbf{d} \in [0, 1]^n$$

Koenker & D'Orey (1994) [A Remark on Algorithm AS 229: Computing Dual Regression Quantiles and Regression Rank Scores](#) suggest to use the **simplex method** (default method in R)

Portnoy & Koenker (1997) [The Gaussian hare and the Laplacian tortoise](#) suggest to use the **interior point method**

## Simplex Method

The beer problem: we want to produce beer, either blonde, or brown

	$\left\{ \begin{array}{l} \text{barley : } 14\text{kg} \\ \text{corn : } 2\text{kg} \\ \text{price : } 30\text{\euro} \end{array} \right.$		$\left\{ \begin{array}{l} \text{barley : } 10\text{kg} \\ \text{corn : } 5\text{kg} \\ \text{price : } 40\text{\euro} \end{array} \right.$		$\left\{ \begin{array}{l} \text{barley : } 280\text{kg} \\ \text{corn : } 100\text{kg} \end{array} \right.$
---	--	---	--	---	---

Admissible sets :

$$10q_{\text{brown}} + 14q_{\text{blond}} \leq 280 \quad (10x_1 + 14x_2 \leq 280)$$

$$2q_{\text{brown}} + 5q_{\text{blond}} \leq 100 \quad (2x_1 + 5x_2 \leq 100)$$

What should we produce to maximize the profit ?

$$\max \{40q_{\text{brown}} + 30q_{\text{blond}}\} \quad (\max \{40x_1 + 30x_2\})$$

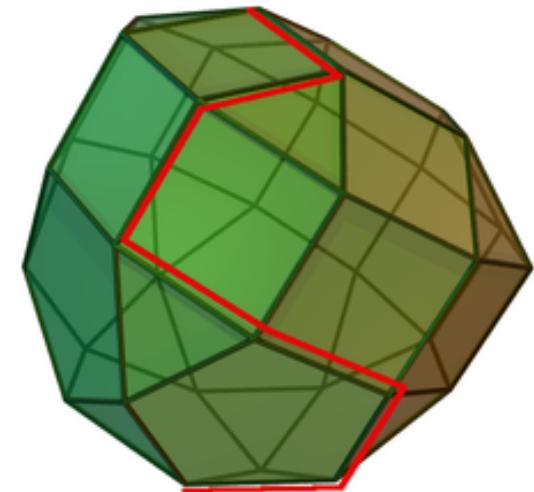
## Simplex Method

First step: enlarge the space,  $10x_1 + 14x_2 \leq 280$  becomes  $10x_1 + 14x_2 - u_1 = 280$  (so called slack variables)

$$\begin{aligned} & \max \{40x_1 + 30x_2\} \\ \text{s.t. } & 10x_1 + 14x_2 + u_1 = 280 \\ & 2x_1 + 5x_2 + u_2 = 100 \\ & x_1, x_2, u_1, u_2 \geq 0 \end{aligned}$$

summarized in the following table, see [wikibook](#)

	$x_1$	$x_2$	$u_1$	$u_2$	
(1)	10	14	1	0	280
(2)	2	5	0	1	100
max	40	30	0	0	



## Interior Point Method

See Vanderbei *et al.* (1986) [A modification of Karmarkar's linear programming algorithm](#) for a presentation of the algorithm, Potra & Wright (2000) [Interior-point methods](#) for a general survey, and Meketon (1986) [Least absolute value regression](#) for an application of the algorithm in the context of median regression.

Running time is of order  $n^{1+\delta}k^3$  for some  $\delta > 0$  and  $k = \dim(\beta)$  (it is  $(n + k)k^2$  for OLS, see [wikipedia](#)).

## Quantile Regression Estimators

OLS estimator  $\hat{\beta}^{\text{ols}}$  is solution of

$$\hat{\beta}^{\text{ols}} = \operatorname{argmin} \left\{ \mathbb{E} [(\mathbb{E}[Y|\mathbf{X}=\mathbf{x}] - \mathbf{x}^\top \beta)^2] \right\}$$

and Angrist, Chernozhukov & Fernandez-Val (2006) [Quantile Regression under Misspecification](#) proved that

$$\hat{\beta}_\tau = \operatorname{argmin} \left\{ \mathbb{E} [\omega_\tau(\beta) (Q_\tau[Y|\mathbf{X}=\mathbf{x}] - \mathbf{x}^\top \beta)^2] \right\}$$

(under weak conditions) where

$$\omega_\tau(\beta) = \int_0^1 (1-u) f_{y|\mathbf{x}}(u \mathbf{x}^\top \beta + (1-u) Q_\tau[Y|\mathbf{X}=\mathbf{x}]) du$$

$\hat{\beta}_\tau$  is the best weighted mean square approximation of the true quantile function, where the weights depend on an average of the conditional density of  $Y$  over  $\mathbf{x}^\top \beta$  and the true quantile regression function.

## Assumptions to get Consistency of Quantile Regression Estimators

As always, we need some assumptions to have consistency of estimators.

- observations  $(Y_i, \mathbf{X}_i)$  must (conditionnaly) i.i.d.
- regressors must have a bounded second moment,  $\mathbb{E}[\|\mathbf{X}_i\|^2] < \infty$
- error terms  $\varepsilon$  are continuously distributed given  $\mathbf{X}_i$ , centered in the sense that their median should be 0,

$$\int_{-\infty}^0 f_\varepsilon(\epsilon) d\epsilon = \frac{1}{2}.$$

- “local identification” property :  $[f_\varepsilon(0) \mathbf{X} \mathbf{X}^\top]$  is positive definite

## Quantile Regression Estimators

Under those weak conditions,  $\hat{\beta}_\tau$  is asymptotically normal:

$$\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau(1-\tau)D_\tau^{-1}\Omega_x D_\tau^{-1}),$$

where

$$D_\tau = \mathbb{E}[f_\varepsilon(0)\mathbf{X}\mathbf{X}^\top] \text{ and } \Omega_x = \mathbb{E}[\mathbf{X}^\top\mathbf{X}].$$

hence, the asymptotic variance of  $\hat{\beta}$  is

$$\widehat{\text{Var}}[\hat{\beta}_\tau] = \frac{\tau(1-\tau)}{[\hat{f}_\varepsilon(0)]^2} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^\top \mathbf{x}_i \right)^{-1}$$

where  $\hat{f}_\varepsilon(0)$  is estimated using (e.g.) an histogram, as suggested in Powell (1991)

**Estimation of monotonic regression models under quantile restrictions**, since

$$D_\tau = \lim_{h \downarrow 0} \mathbb{E} \left( \frac{\mathbf{1}(|\varepsilon| \leq h)}{2h} \mathbf{X}\mathbf{X}^\top \right) \sim \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}(|\varepsilon_i| \leq h) \mathbf{x}_i \mathbf{x}_i^\top = \hat{D}_\tau.$$

## Quantile Regression Estimators

There is no first order condition, in the sense  $\partial V_n(\beta, \tau)/\partial \beta = \mathbf{0}$  where

$$V_n(\beta, \tau) = \sum_{i=1}^n \mathcal{R}_\tau^q(y_i - \mathbf{x}_i^\top \beta)$$

There is an asymptotic first order condition,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \psi_\tau(y_i - \mathbf{x}_i^\top \beta) = \mathcal{O}(1), \text{ as } n \rightarrow \infty,$$

where  $\psi_\tau(\cdot) = \mathbf{1}(\cdot < 0) - \tau$ , see Huber (1967) [The behavior of maximum likelihood estimates under nonstandard conditions](#).

One can also define a Wald test, a Likelihood Ratio test, etc.

## Quantile Regression Estimators

Then the confidence interval of level  $1 - \alpha$  is then

$$\left[ \hat{\beta}_\tau \pm z_{1-\alpha/2} \sqrt{\text{Var}[\hat{\beta}_\tau]} \right]$$

An alternative is to use a bootstrap strategy (see #2)

- generate a sample  $(y_i^{(b)}, \mathbf{x}_i^{(b)})$  from  $(y_i, \mathbf{x}_i)$
- estimate  $\hat{\beta}_\tau^{(b)}$  by

$$\hat{\beta}_\tau^{(b)} = \operatorname{argmin} \left\{ \mathcal{R}_\tau^q(y_i^{(b)} - \mathbf{x}_i^{(b)\top} \beta) \right\}$$

- set  $\hat{\text{Var}}^\star[\hat{\beta}_\tau] = \frac{1}{B} \sum_{b=1}^B (\hat{\beta}_\tau^{(b)} - \hat{\beta}_\tau)^2$

For confidence intervals, we can either use Gaussian-type confidence intervals, or empirical quantiles from bootstrap estimates.

## Quantile Regression Estimators

If  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)$ , one can prove that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} - \boldsymbol{\beta}_{\boldsymbol{\tau}}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\tau}}),$$

where  $\Sigma_{\boldsymbol{\tau}}$  is a block matrix, with

$$\Sigma_{\tau_i, \tau_j} = (\min\{\tau_i, \tau_j\} - \tau_i \tau_j) D_{\tau_i}^{-1} \Omega_x D_{\tau_j}^{-1}$$

see Kocherginsky *et al.* (2005) [Practical Confidence Intervals for Regression Quantiles](#) for more details.

## Quantile Regression: Transformations

Scale equivariance

For any  $a > 0$  and  $\tau \in [0, 1]$

$$\hat{\beta}_\tau(aY, \mathbf{X}) = a\hat{\beta}_\tau(Y, \mathbf{X}) \text{ and } \hat{\beta}_\tau(-aY, \mathbf{X}) = -a\hat{\beta}_{1-\tau}(Y, \mathbf{X})$$

Equivariance to reparameterization of design

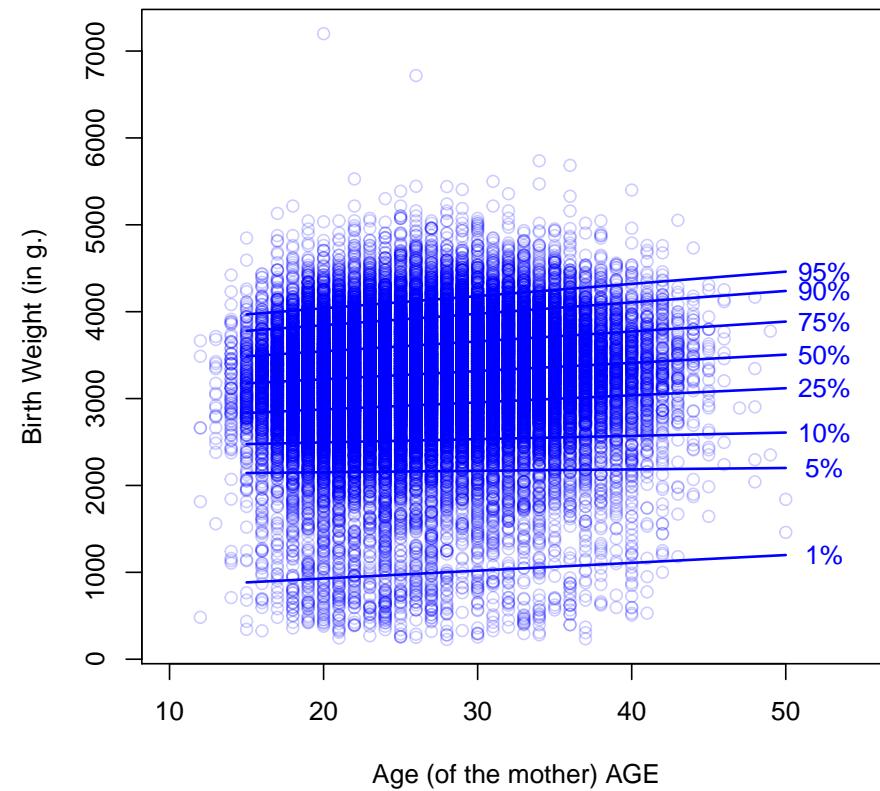
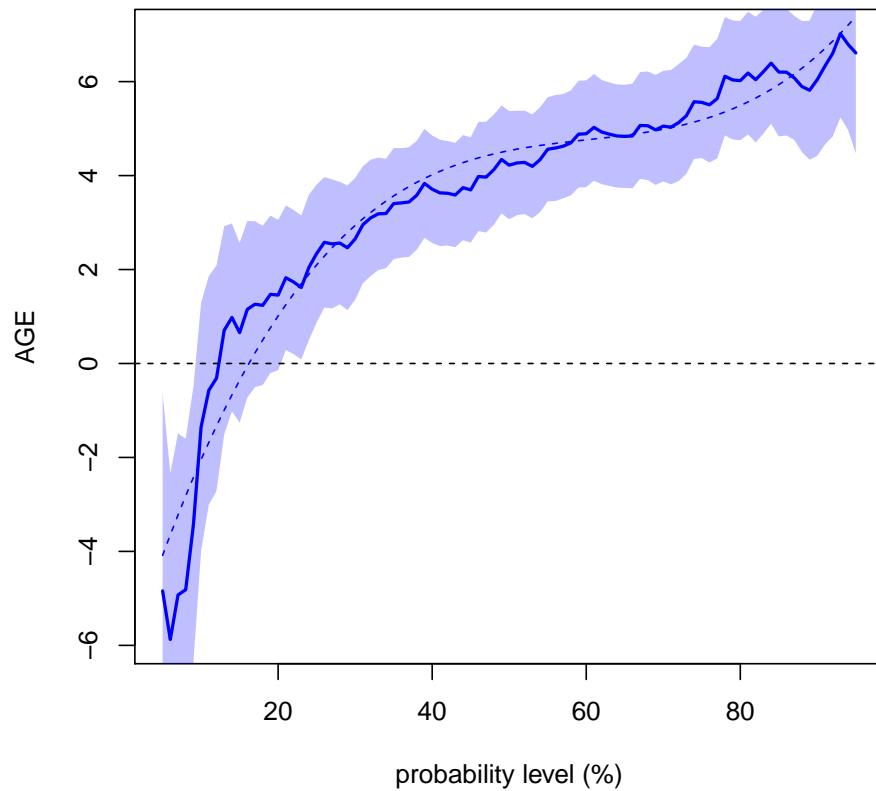
Let  $\mathbf{A}$  be any  $p \times p$  nonsingular matrix and  $\tau \in [0, 1]$

$$\hat{\beta}_\tau(Y, \mathbf{X}\mathbf{A}) = \mathbf{A}^{-1}\hat{\beta}_\tau(Y, \mathbf{X})$$

## Visualization, $\tau \mapsto \hat{\beta}_\tau$

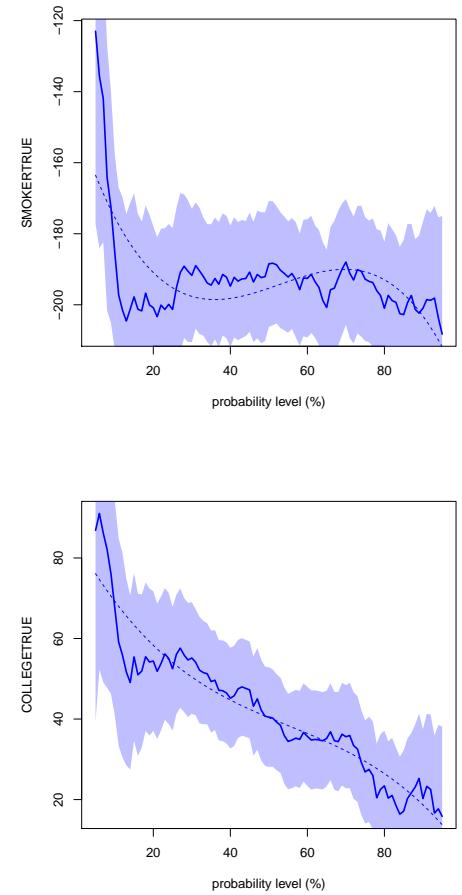
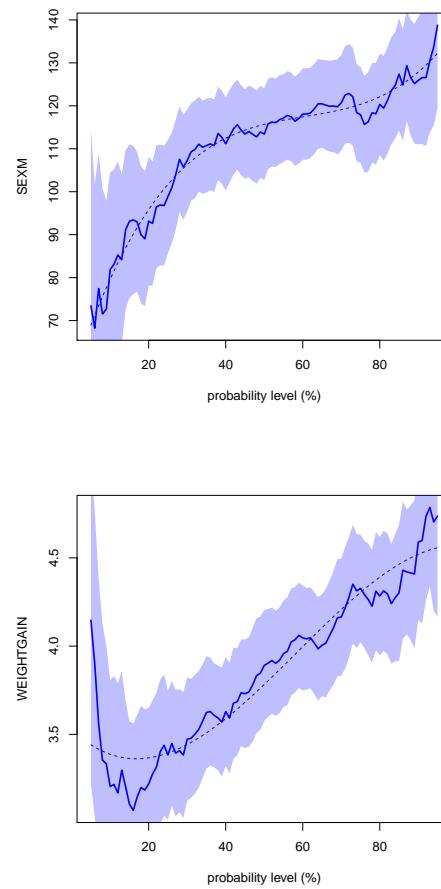
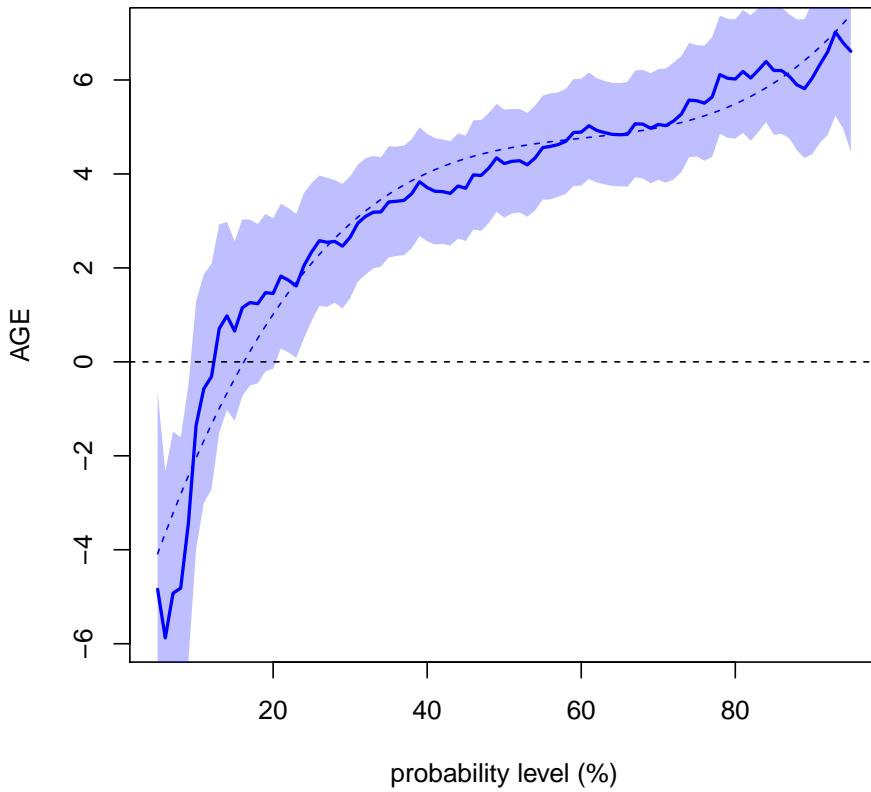
See Abrevaya (2001) [The effects of demographics and maternal behavior...](#)

```
1 > base=read.table("http://freakonometrics.free.fr/natality2005.txt")
```



## Visualization, $\tau \mapsto \hat{\beta}_\tau$

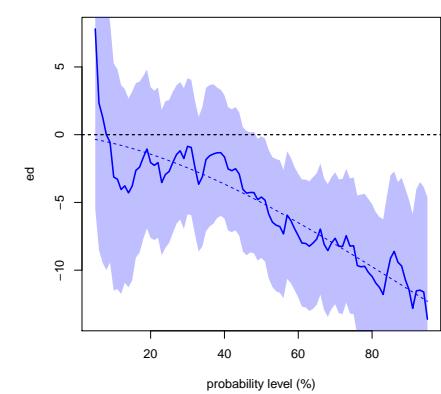
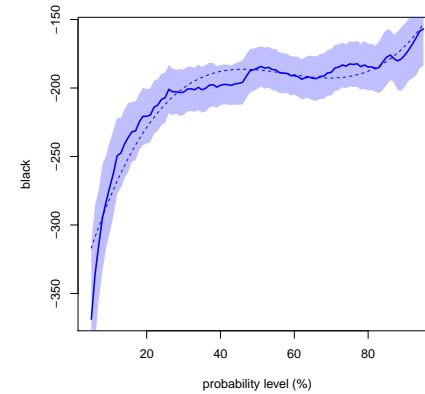
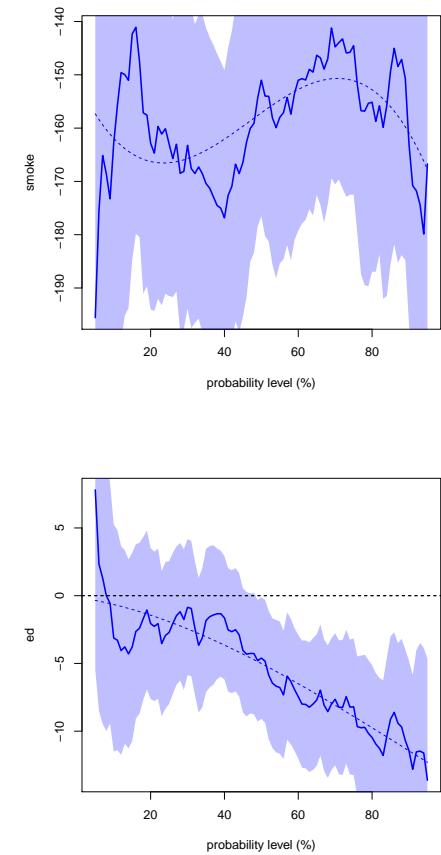
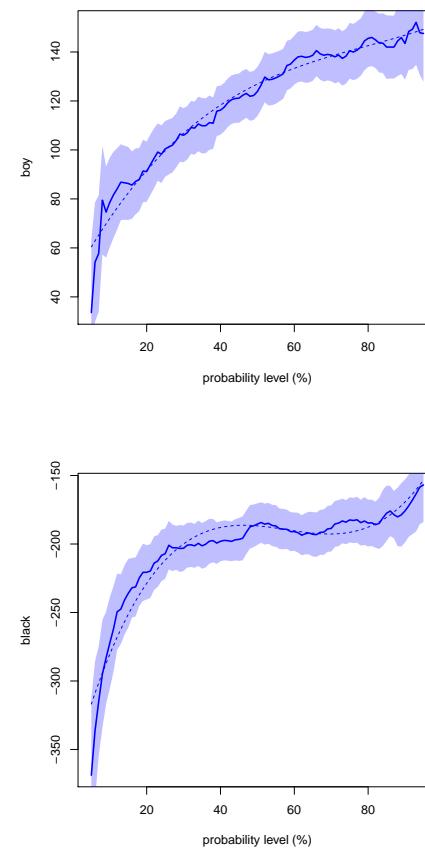
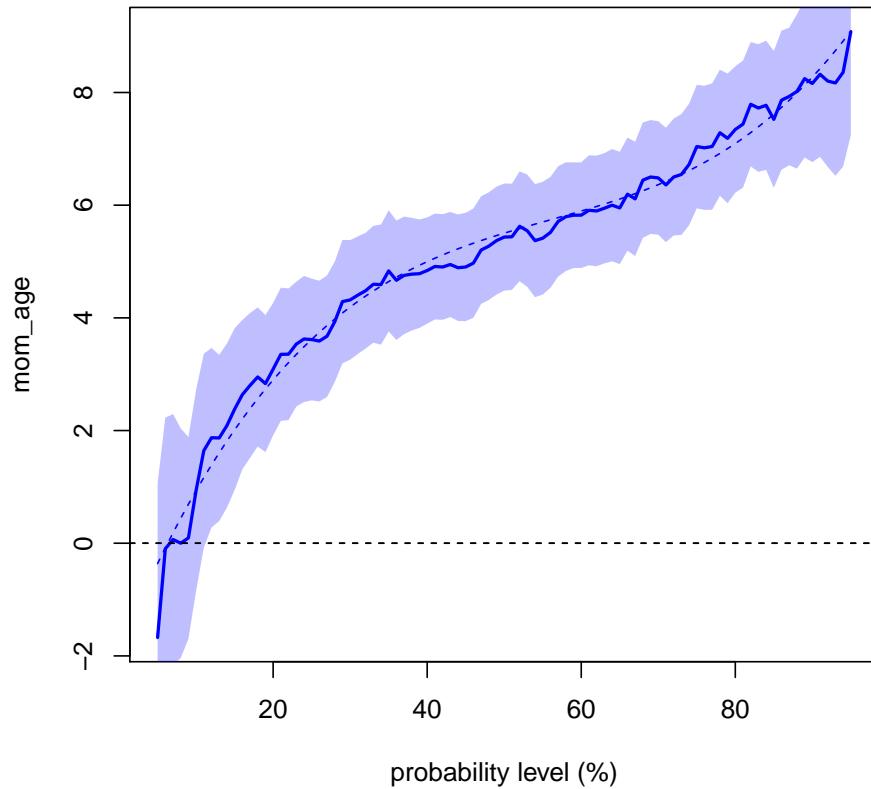
See Abreveya (2001) **The effects of demographics and maternal behavior on the distribution of birth outcomes**



## Visualization, $\tau \mapsto \hat{\beta}_\tau$

See Abrevaya (2001) [The effects of demographics and maternal behavior...](#)

```
1 > base=read.table("http://freakonometrics.free.fr/BWeight.csv")
```

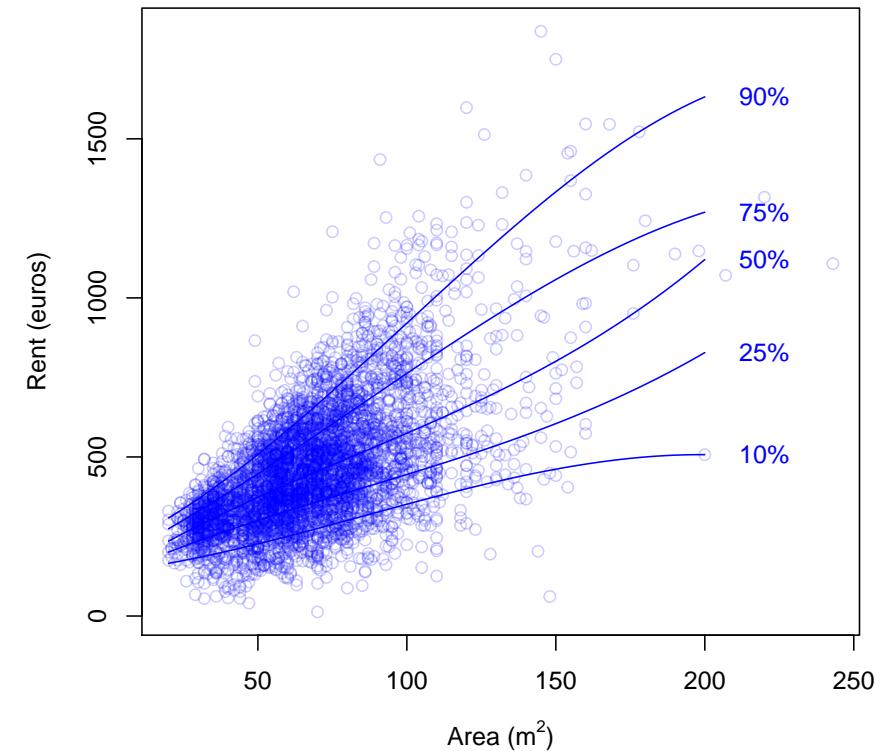
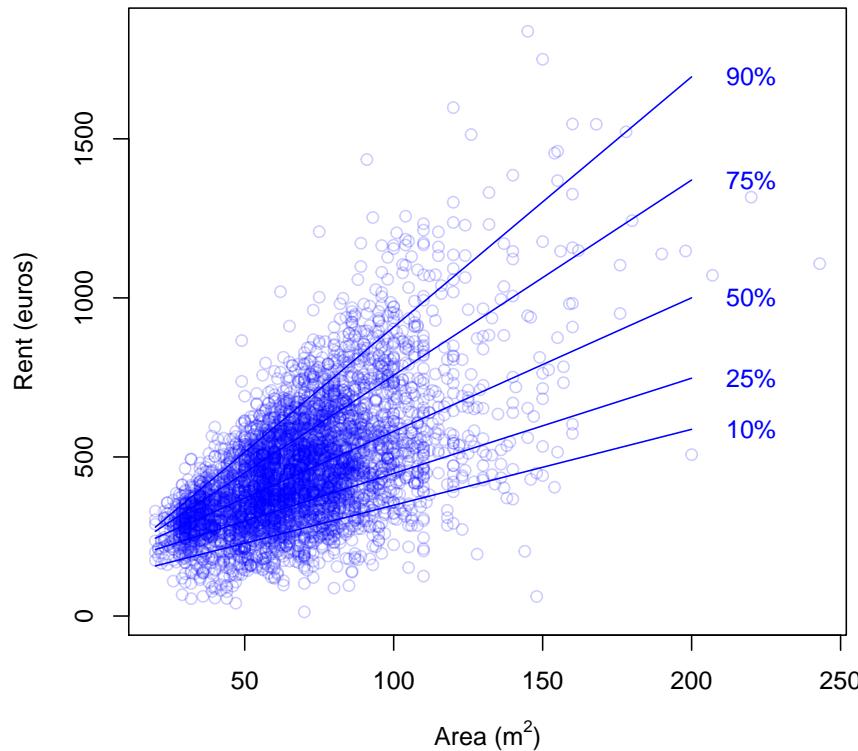


## Quantile Regression, with Non-Linear Effects

Rents in Munich, as a function of the area, from Fahrmeir *et al.* (2013)

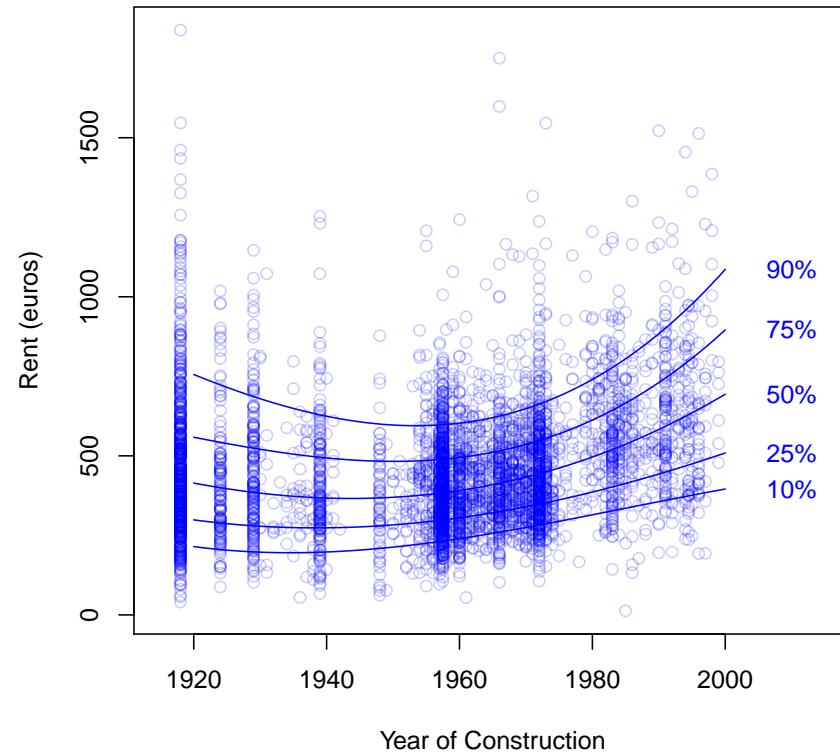
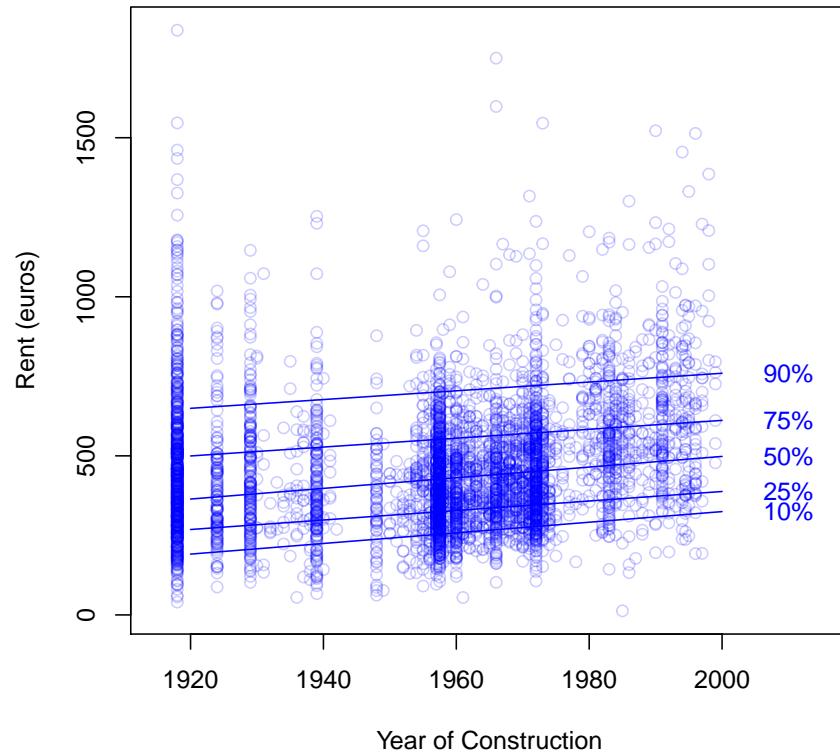
Regression: Models, Methods and Applications

```
1 > base=read.table("http://freakonometrics.free.fr/rent98_00.txt")
```



## Quantile Regression, with Non-Linear Effects

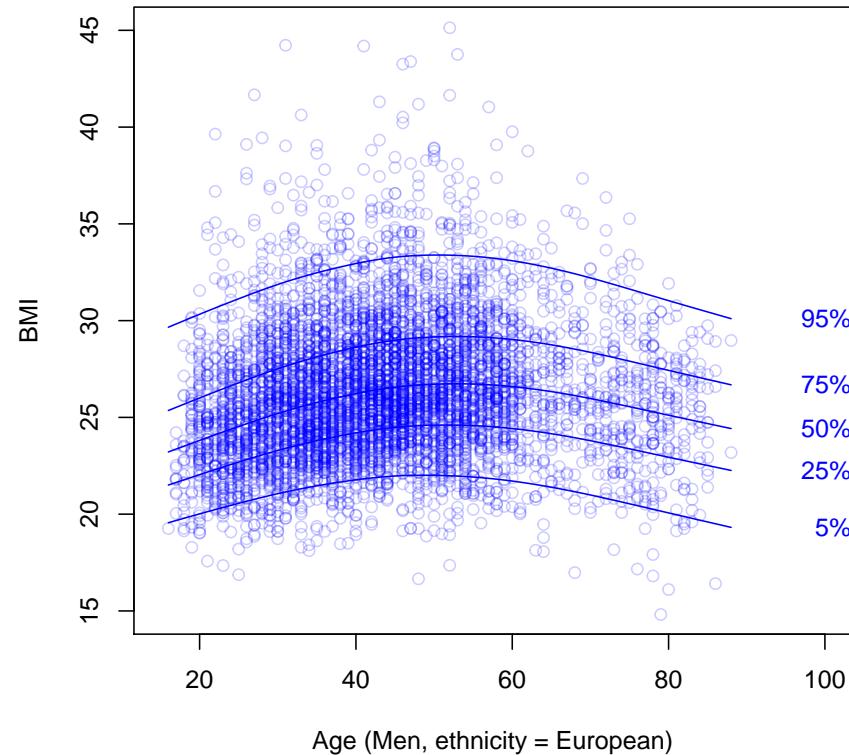
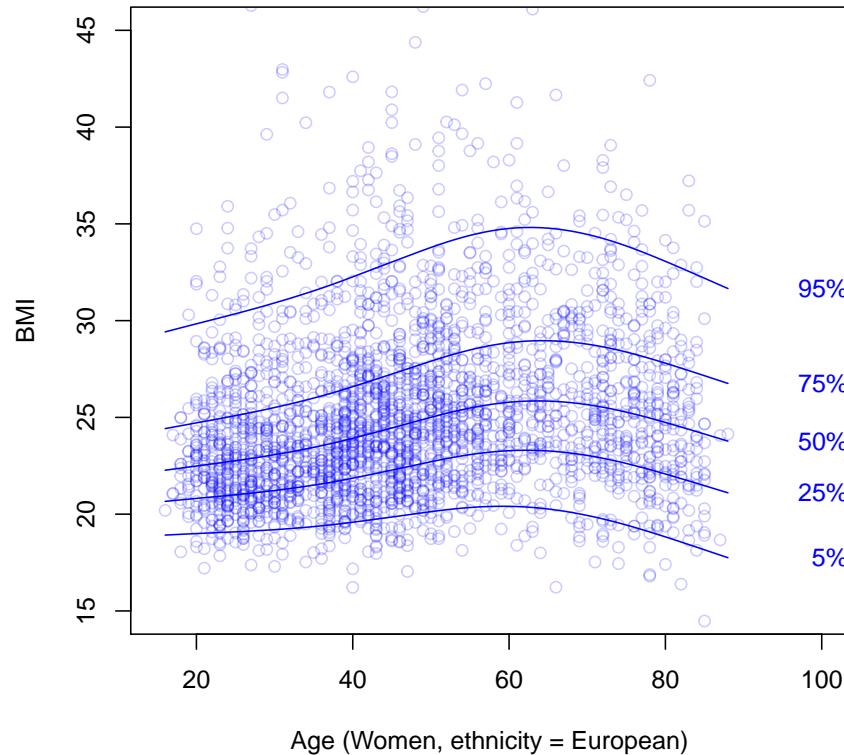
Rents in Munich, as a function of the year of construction, from Fahrmeir *et al.* (2013) *Regression: Models, Methods and Applications*



## Quantile Regression, with Non-Linear Effects

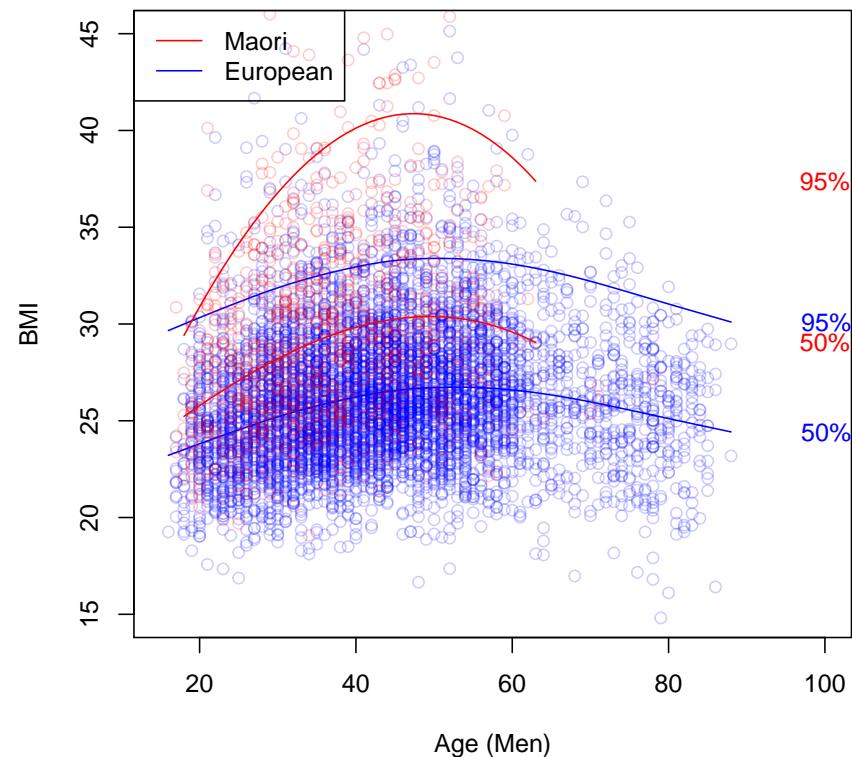
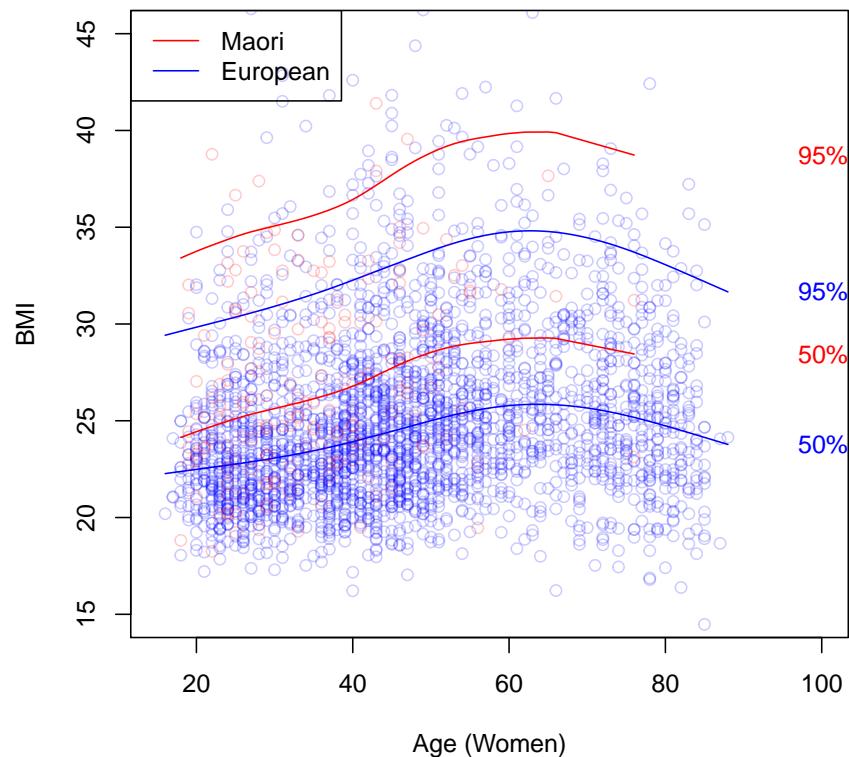
BMI as a function of the age, in New-Zealand, from Yee (2015) [Vector Generalized Linear and Additive Models](#), for Women and Men

```
1 > library(VGAMdata); data(xs.nz)
```



## Quantile Regression, with Non-Linear Effects

BMI as a function of the age, in New-Zealand, from Yee (2015) [Vector Generalized Linear and Additive Models](#), for Women and Men



## Quantile Regression, with Non-Linear Effects

One can consider some local polynomial quantile regression, e.g.

$$\min \left\{ \sum_{i=1}^n \omega_i(\boldsymbol{x}) \mathcal{R}_\tau^q(y_i - \beta_0 - (\boldsymbol{x}_i - \boldsymbol{x})^\top \boldsymbol{\beta}_1) \right\}$$

for some weights  $\omega_i(\boldsymbol{x}) = H^{-1} K(H^{-1}(\boldsymbol{x}_i - \boldsymbol{x}))$ , see Fan, Hu & Truong (1994)

Robust Non-Parametric Function Estimation.

## Asymmetric Maximum Likelihood Estimation

Introduced by Efron (1991) **Regression percentiles using asymmetric squared error loss**. Consider a linear model,  $y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$ . Let

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n Q_\omega(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}), \text{ where } Q_\omega(\epsilon) = \begin{cases} \epsilon^2 & \text{if } \epsilon \leq 0 \\ w\epsilon^2 & \text{if } \epsilon > 0 \end{cases} \quad \text{where } w = \frac{\omega}{1-\omega}$$

One might consider  $\omega_\alpha = 1 + \frac{z_\alpha}{\varphi(z_\alpha) + (1-\alpha)z_\alpha}$  where  $z_\alpha = \Phi^{-1}(\alpha)$ .

Efron (1992) **Poisson overdispersion estimates based on the method of asymmetric maximum likelihood** introduced asymmetric maximum likelihood (AML) estimation, considering

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n Q_\omega(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}), \text{ where } Q_\omega(\epsilon) = \begin{cases} D(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}) & \text{if } y_i \leq \mathbf{x}_i^\top \boldsymbol{\beta} \\ wD(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}) & \text{if } y_i > \mathbf{x}_i^\top \boldsymbol{\beta} \end{cases}$$

where  $D(\cdot, \cdot)$  is the deviance. Estimation is based on Newton-Raphson (gradient descent).

## Noncrossing Solutions

See Bondell *et al.* (2010) **Non-crossing quantile regression curve estimation**.

Consider probabilities  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_q)$  with  $0 < \tau_1 < \dots < \tau_q < 1$ .

Use **parallelism** : add constraints in the optimization problem, such that

$$\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\tau_j} \geq \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\tau_{j-1}} \quad \forall i \in \{1, \dots, n\}, j \in \{2, \dots, q\}.$$

## Quantile Regression on Panel Data

In the context of panel data, consider some fixed effect,  $\alpha_i$  so that

$$y_{i,t} = \mathbf{x}_{i,t}^\top \boldsymbol{\beta}_\tau + \alpha_i + \varepsilon_{i,t} \text{ where } Q_\tau(\varepsilon_{i,t} | \mathbf{X}_i) = 0$$

Canay (2011) [A simple approach to quantile regression for panel data](#) suggests an estimator in two steps,

- use a standard OLS fixed-effect model  $y_{i,t} = \mathbf{x}_{i,t}^\top \boldsymbol{\beta} + \alpha_i + u_{i,t}$ , i.e. consider a within transformation, and derive the fixed effect estimate  $\hat{\boldsymbol{\beta}}$

$$(y_{i,t} - \bar{y}_i) = (\mathbf{x}_{i,t} - \bar{\mathbf{x}}_{i,t})^\top \boldsymbol{\beta} + (u_{i,t} - \bar{u}_i)$$

- estimate fixed effects as  $\hat{\alpha}_i = \frac{1}{T} \sum_{t=1}^T (y_{i,t} - \mathbf{x}_{i,t}^\top \hat{\boldsymbol{\beta}})$
- finally, run a standard quantile regression of  $y_{i,t} - \hat{\alpha}_i$  on  $\mathbf{x}_{i,t}$ 's.

See `rqpd` package.

## Quantile Regression with Fixed Effects (QRFE)

In a panel linear regression model,  $y_{i,t} = \mathbf{x}_{i,t}^\top \boldsymbol{\beta} + u_i + \varepsilon_{i,t}$ ,

where  $u$  is an unobserved individual specific effect.

In a fixed effects models,  $u$  is treated as a parameter. Quantile Regression is

$$\min_{\boldsymbol{\beta}, \mathbf{u}} \left\{ \sum_{i,t} \mathcal{R}_\alpha^q (y_{i,t} - [\mathbf{x}_{i,t}^\top \boldsymbol{\beta} + u_i]) \right\}$$

Consider **Penalized QRFE**, as in Koenker & Bilias (2001) **Quantile regression for duration data**,

$$\min_{\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_\kappa, \mathbf{u}} \left\{ \sum_{\mathbf{k}, i, t} \omega_{\mathbf{k}} \mathcal{R}_{\alpha_{\mathbf{k}}}^q (y_{i,t} - [\mathbf{x}_{i,t}^\top \boldsymbol{\beta}_{\mathbf{k}} + u_i]) + \lambda \sum_i |u_i| \right\}$$

where  $\omega_{\mathbf{k}}$  is a relative weight associated with quantile of level  $\alpha_{\mathbf{k}}$ .

## Quantile Regression with Random Effects (QRRE)

Assume here that  $y_{i,t} = \mathbf{x}_{i,t}^\top \boldsymbol{\beta} + \underbrace{u_i + \varepsilon_{i,t}}_{=\eta_{i,t}}.$

Quantile Regression Random Effect (QRRE) yields solving

$$\min_{\boldsymbol{\beta}} \left\{ \sum_{i,t} \mathcal{R}_\alpha^q(y_{i,t} - \mathbf{x}_{i,t}^\top \boldsymbol{\beta}) \right\}$$

which is a weighted asymmetric least square deviation estimator.

Let  $\Sigma = [\sigma_{s,t}(\alpha)]$  denote the matrix

$$\sigma_{ts}(\alpha) = \begin{cases} \alpha(1-\alpha) & \text{if } t = s \\ \mathbb{E}[\mathbf{1}\{\varepsilon_{it}(\alpha) < 0, \varepsilon_{is}(\alpha) < 0\}] - \alpha^2 & \text{if } t \neq s \end{cases}$$

If  $(nT)^{-1} \mathbf{X}^\top \{\mathbb{I}_n \otimes \Sigma_{T \times T}(\alpha)\} \mathbf{X} \rightarrow \mathbf{D}_0$  as  $n \rightarrow \infty$  and  $(nT)^{-1} \mathbf{X}^\top \Omega_f \mathbf{X} = \mathbf{D}_1$ , then

$$\sqrt{nT} \left( \hat{\boldsymbol{\beta}}^Q(\alpha) - \boldsymbol{\beta}^Q(\alpha) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \mathbf{D}_1^{-1} \mathbf{D}_0 \mathbf{D}_0^\top \mathbf{D}_1^{-1} \right).$$

## Quantile Treatment Effects

Doksum (1974) [Empirical Probability Plots and Statistical Inference for Nonlinear Models](#) introduced QTE - Quantile Treatment Effect - when a person might have two  $Y$ 's : either  $Y_0$  (without treatment,  $D = 0$ ) or  $Y_1$  (with treatment,  $D = 1$ ),

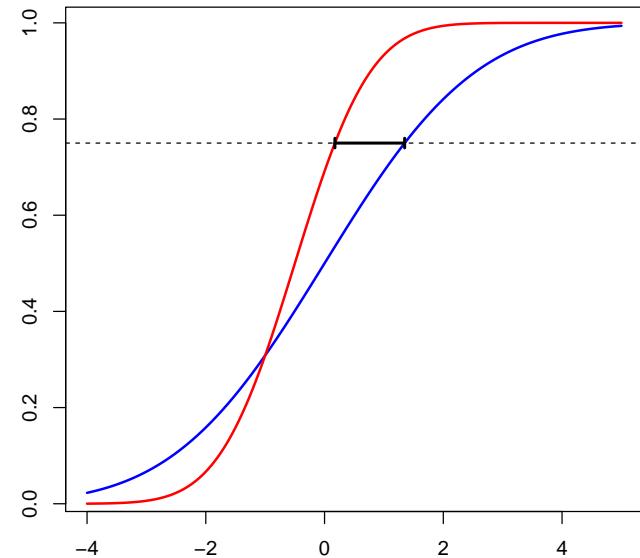
$$\delta_\tau = Q_{Y_1}(\tau) - Q_{Y_0}(\tau)$$

which can be studied on the context of covariates.

Run a quantile regression of  $y$  on  $(d, \mathbf{x})$ ,

$$y = \beta_0 + \delta d + \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i \text{ : shifting effect}$$

$$y = \beta_0 + \mathbf{x}_i^\top [\boldsymbol{\beta} + \delta d] + \varepsilon_i \text{ : scaling effect}$$



## Quantile Regression for Time Series

Consider some GARCH(1,1) financial time series,

$$y_t = \sigma_t \varepsilon_t \text{ where } \sigma_t = \alpha_0 + \alpha_1 \cdot |y_{t-1}| + \beta_1 \sigma_{t-1}.$$

The quantile function conditional on the past -  $\mathcal{F}_{t-1} = \underline{Y}_{t-1}$  - is

$$Q_{y|\mathcal{F}_{t-1}}(\tau) = \underbrace{\alpha_0 F_\varepsilon^{-1}(\tau)}_{\tilde{\alpha}_0} + \underbrace{\alpha_1 F_\varepsilon^{-1}(\tau) \cdot |y_{t-1}|}_{\tilde{\alpha}_1} + \beta_1 Q_{y|\mathcal{F}_{t-2}}(\tau)$$

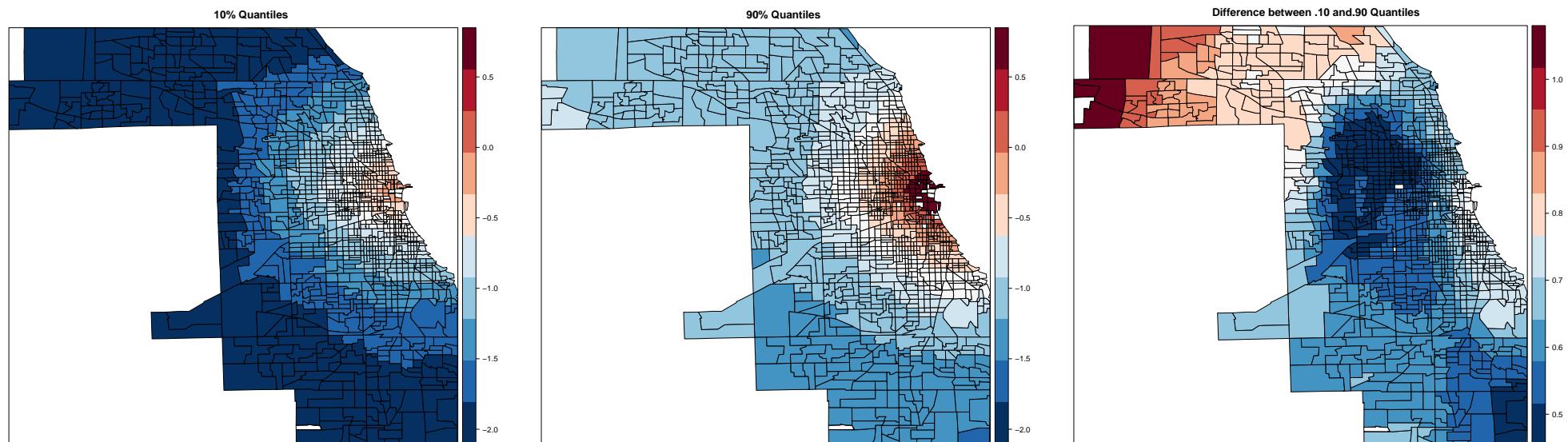
i.e. the conditional quantile has a GARCH(1,1) form, see **Conditional Autoregressive Value-at-Risk**, see Manganelli & Engle (2004) **CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles**

## Quantile Regression for Spatial Data

```

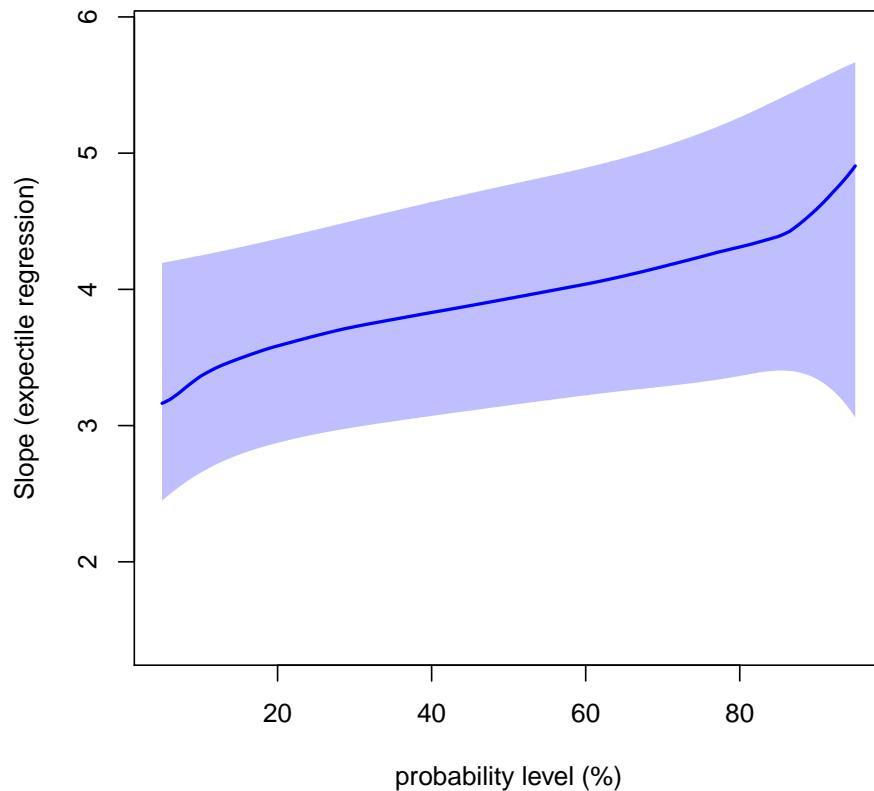
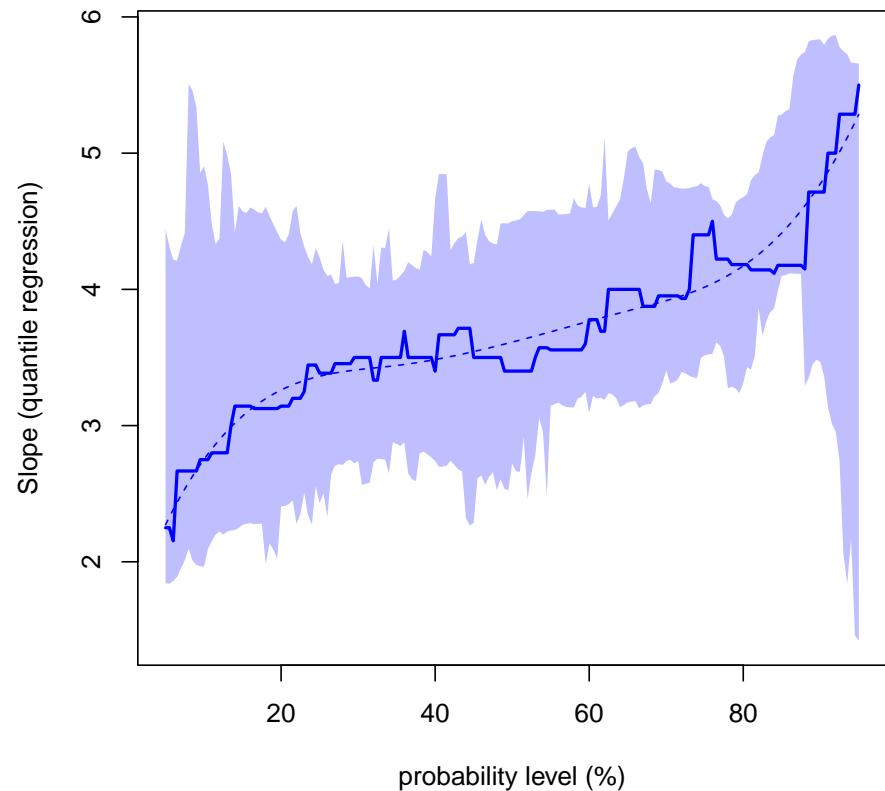
1 > library(McSpatial)
2 > data(cookdata)
3 > fit <- qregcpar(LNFAR~DCBD, nonpar=~LATITUDE+LONGITUDE, taumat=c
  (.10,.90), kern="bisq", window=.30, distance="LATLONG", data=
  cookdata)

```



## Expectile Regression

Quantile regression vs. Expectile regression, on the same dataset (`cars`)



see Koenker (2014) [Living Beyond our Means](#) for a comparison quantiles-expectiles

## Expectile Regression

Solve here  $\min_{\beta} \left\{ \sum_{i=1}^n \mathcal{R}_\tau^e(y_i - \mathbf{x}_i^\top \beta) \right\}$  where  $\mathcal{R}_\tau^e(u) = u^2 \cdot (\tau - \mathbf{1}(u < 0))$

*“this estimator can be interpreted as a maximum likelihood estimator when the disturbances arise from a normal distribution with unequal weight placed on positive and negative disturbances”* Aigner, Amemiya & Poirier (1976)

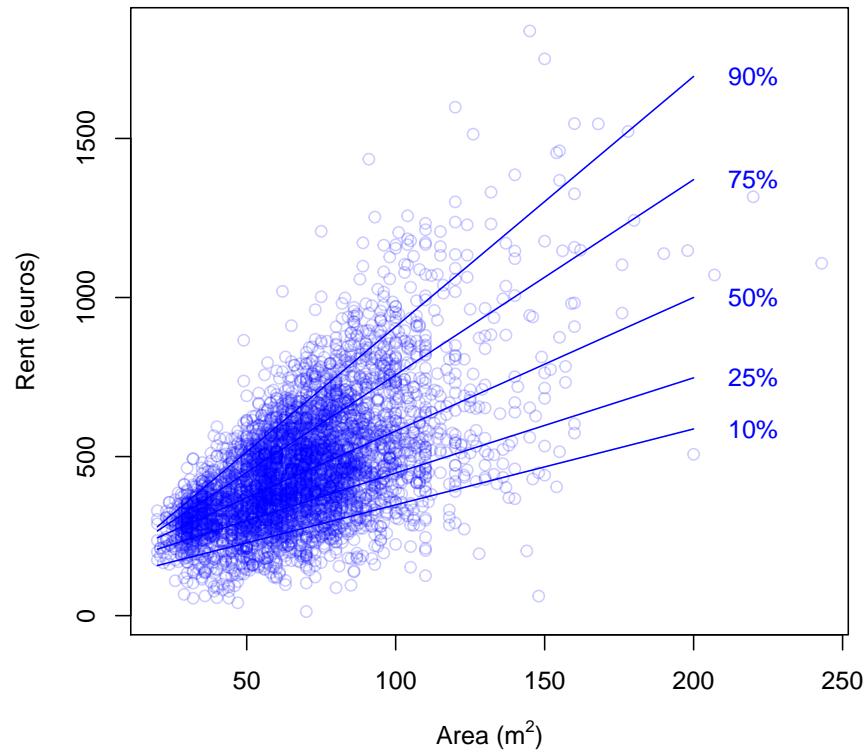
### Formulation and Estimation of Stochastic Frontier Production Function Models.

See Holzmann & Klar (2016) [Expectile Asymptotics](#) for statistical properties.

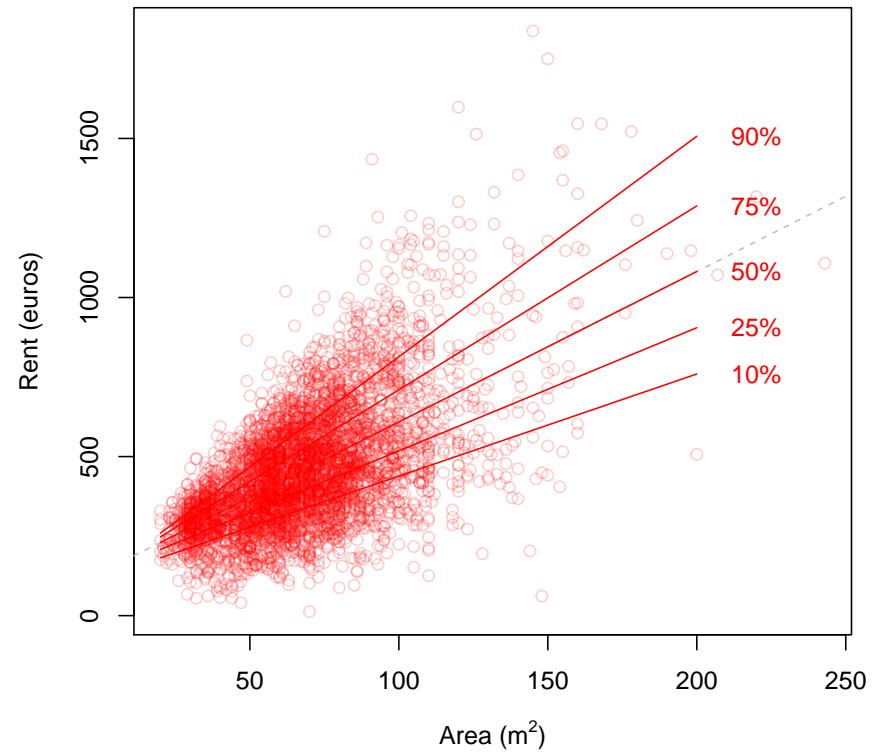
Expectiles can (also) be related to Breckling & Chambers (1988) [M-Quantiles](#).

Comparison quantile regression and expectile regression, see Schulze-Waltrup *et al.* (2014) [Expectile and quantile regression - David and Goliath?](#)

## Expectile Regression, with Linear Effects



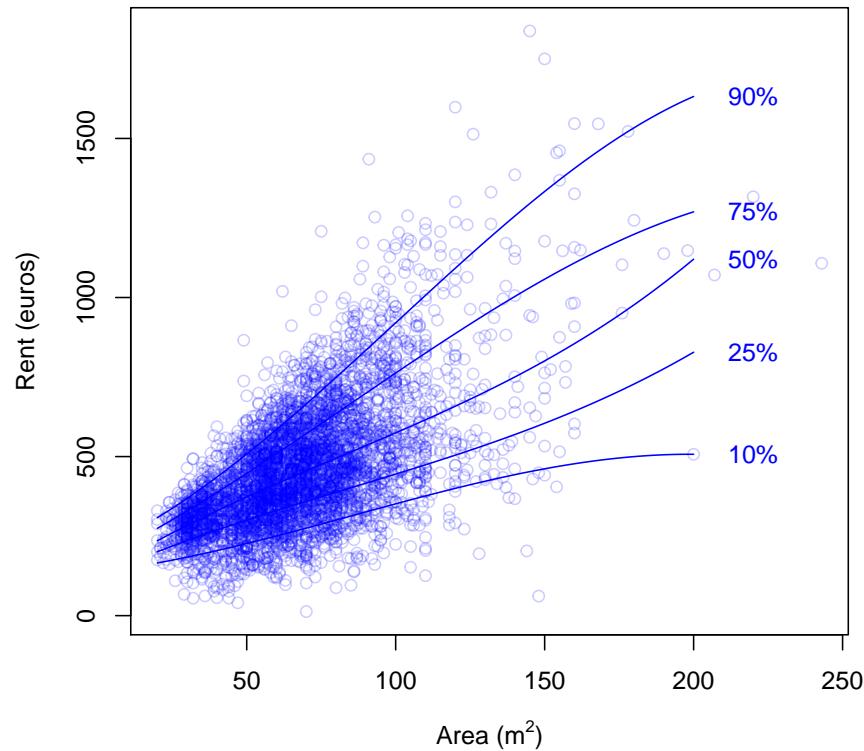
Quantile Regressions



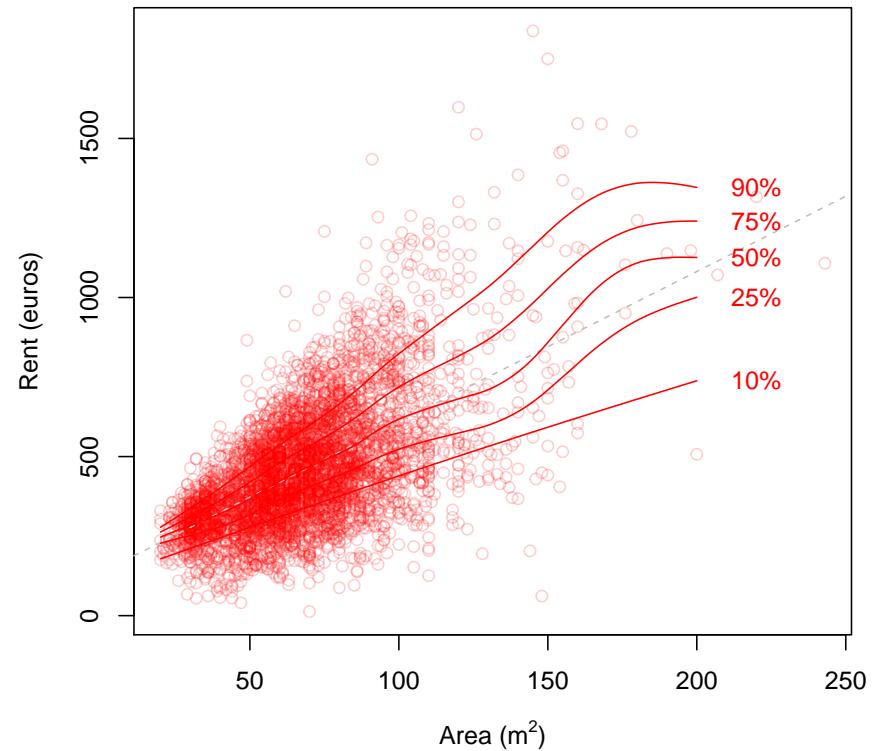
Expectile Regressions

## Expectile Regression, with Non-Linear Effects

See Zhang (1994) Nonparametric regression expectiles



Quantile Regressions



Expectile Regressions

## Expectile Regression, with Random Effects (ERRE)

Quantile Regression Random Effect (QRRE) yields solving

$$\min_{\beta} \left\{ \sum_{i,t} \mathcal{R}_\alpha^e(y_{i,t} - \mathbf{x}_{i,t}^\top \beta) \right\}$$

One can prove that

$$\hat{\beta}^e(\tau) = \left( \sum_{i=1}^n \sum_{t=1}^T \hat{\omega}_{i,t}(\tau) \mathbf{x}_{it} \mathbf{x}_{it}^\top \right)^{-1} \left( \sum_{i=1}^n \sum_{t=1}^T \hat{\omega}_{i,t}(\tau) \mathbf{x}_{it} y_{it} \right),$$

where  $\hat{\omega}_{it}(\tau) = |\tau - \mathbf{1}(y_{it} < \mathbf{x}_{it}^\top \hat{\beta}^e(\tau))|$ .

## Expectile Regression with Random Effects (ERRE)

If  $W = \text{diag}(\omega_{11}(\tau), \dots, \omega_{nT}(\tau))$ , set

$$\bar{W} = \mathbb{E}(W), H = \mathbf{X}^\top \bar{W} \mathbf{X} \text{ and } \Sigma = \mathbf{X}^\top \mathbb{E}(W \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^\top W) \mathbf{X}.$$

and then

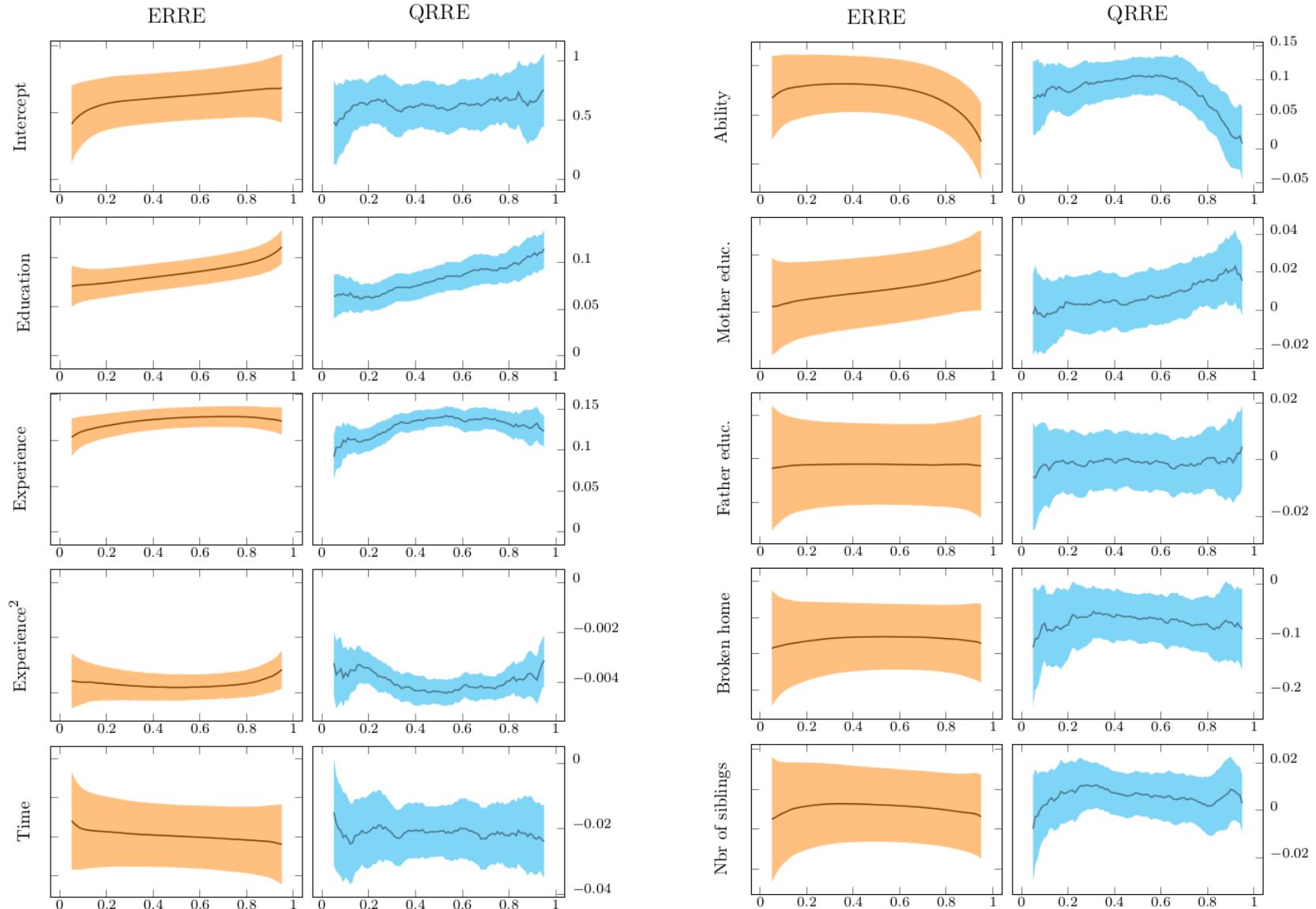
$$\sqrt{nT} \{ \hat{\boldsymbol{\beta}}^e(\tau) - \boldsymbol{\beta}^e(\tau) \} \xrightarrow{\mathcal{L}} \mathcal{N}(0, H^{-1} \Sigma H^{-1}),$$

see Barry *et al.* (2016) Quantile and Expectile Regression for random effects model.

See, for expectile regressions, with R,

```
1 > library(expectreg)
2 > fit <- expectreg.ls(rent_euro ~ area, data=munich, expectiles=.75)
3 > fit <- expectreg.ls(rent_euro ~ rb(area,"pspline"), data=munich,
   expectiles=.75)
```

## Application to Real Data



## Extensions

The mean of  $Y$  is  $\nu(F_Y) = \int_{-\infty}^{+\infty} y dF_Y(y)$

The quantile of level  $\tau$  for  $Y$  is  $\nu_\tau(F_Y) = F_Y^{-1}(\tau)$

More generally, consider some functional  $\nu(F)$  (Gini or Theil index, entropy, etc),  
see Foresi & Peracchi (1995) [The Conditional Distribution of Excess Returns](#)

Can we estimate  $\nu(F_{Y|\mathbf{x}})$  ?

Firpo *et al.* (2009) [Unconditional Quantile Regressions](#) suggested to use [influence function regression](#)

Machado & Mata (2005) [Counterfactual decomposition of changes in wage distributions](#) and Chernozhukov *et al.* (2013) [Inference on counterfactual distributions](#) suggested indirect distribution function.

Influence function of index  $\nu(F)$  at  $y$  is

$$IF(y, \nu, F) = \lim_{\epsilon \downarrow 0} \frac{\nu((1 - \epsilon)F + \epsilon\delta_y) - \nu(F)}{\epsilon}$$

## Extensions

For Gini index, and Generalized Entropy, see Cowell & Flachaire (2007) [Income distribution and inequality measurement: The problem of extreme values](#).

One can also look for the conditional distribution by estimating (uniform) sequence of conditional quantile prediction...