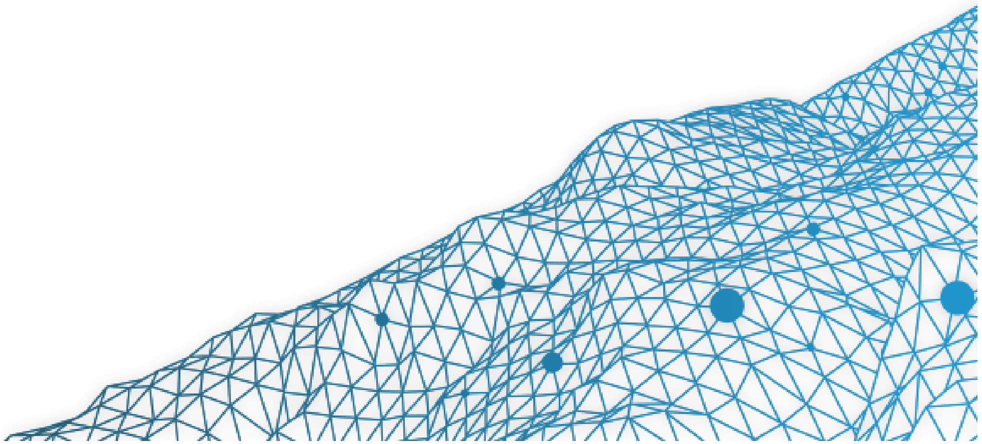
2 Least Squares & Other Loss Functions

Arthur Charpentier (Université du Québec à Montréal)

Machine Learning & Econometrics

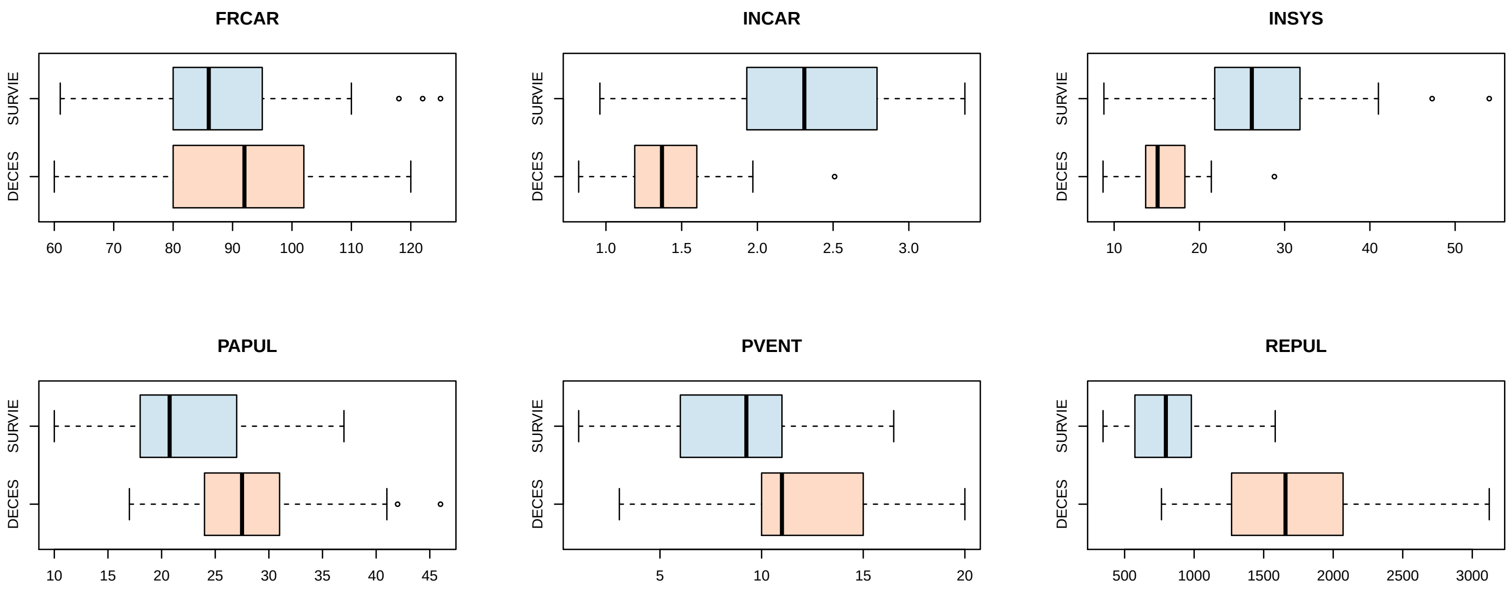
SIDE Summer School - July 2019



Quantiles and Loss Functions

Let Y denote a random variable with cumulative distribution function F, $F(y) = \mathbb{P}[Y \leq y]$. The quantile is

$$Q(u) = \inf\big\{x \in \mathbb{R}, F(x) > u\big\}.$$



Box plot, from Tukey (1977, Exploratory Data Analysis).

Defining halfspace depth

Given $\boldsymbol{y} \in \mathbb{R}^d$, and a direction $\boldsymbol{u} \in \mathbb{R}^d$, define the closed half space

$$H_{\boldsymbol{y},\boldsymbol{u}} = \{\boldsymbol{x} \in \mathbb{R}^d \text{ such that } \boldsymbol{u}'\boldsymbol{x} \leq \boldsymbol{u}'\boldsymbol{y}\}$$

and define depth at point $\boldsymbol{y}$ by

$$\text{depth}(\boldsymbol{y}) = \inf_{\boldsymbol{u},\boldsymbol{u}\neq\boldsymbol{0}} \{\mathbb{P}(H_{\boldsymbol{y},\boldsymbol{u}})\}$$

i.e. the smallest probability of a closed half space containing $\boldsymbol{y}$.

The empirical version is - see Tukey (1975, Mathematics and the picturing of data)

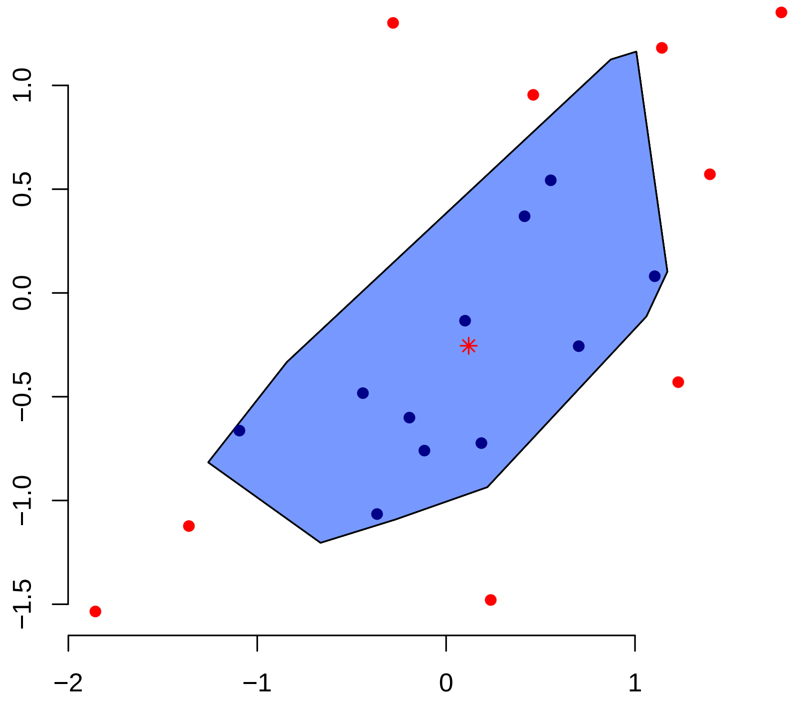
$$\text{depth}(\boldsymbol{y}) = \min_{\boldsymbol{u},\boldsymbol{u}\neq\boldsymbol{0}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\boldsymbol{X}_i \in H_{\boldsymbol{y},\boldsymbol{u}}) \right\}$$

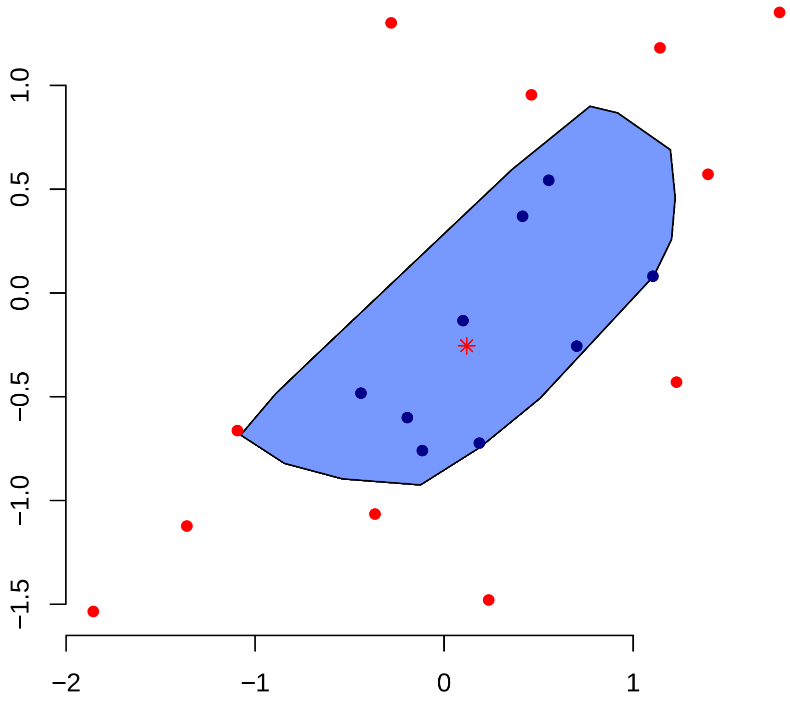
For $\alpha > 0.5$, define the depth set as

$$D_\alpha = \{\boldsymbol{y} \in \mathbb{R} \in \mathbb{R}^d \text{ such that } \geq 1 - \alpha\}.$$

The empirical version is can be related to the bagplot, Rousseeuw *et al.*, 1999.

Empirical sets extremely sentive to the algorithm





where the blue set is the empirical estimation for D_α, $\alpha = 0.5$.

The bagplot tool

The depth function introduced here is the multivariate extension of standard univariate depth measures, e.g.

$$\text{depth}(x) = \min\{F(x), 1 - F(x^-)\}$$

which satisfies $\text{depth}(Q_\alpha) = \min\{\alpha, 1 - \alpha\}$. But one can also consider

$$\text{depth}(x) = 2 \cdot F(x) \cdot [1 - F(x^-)] \text{ or } \text{depth}(x) = 1 - \left|\frac{1}{2} - F(x)\right|.$$

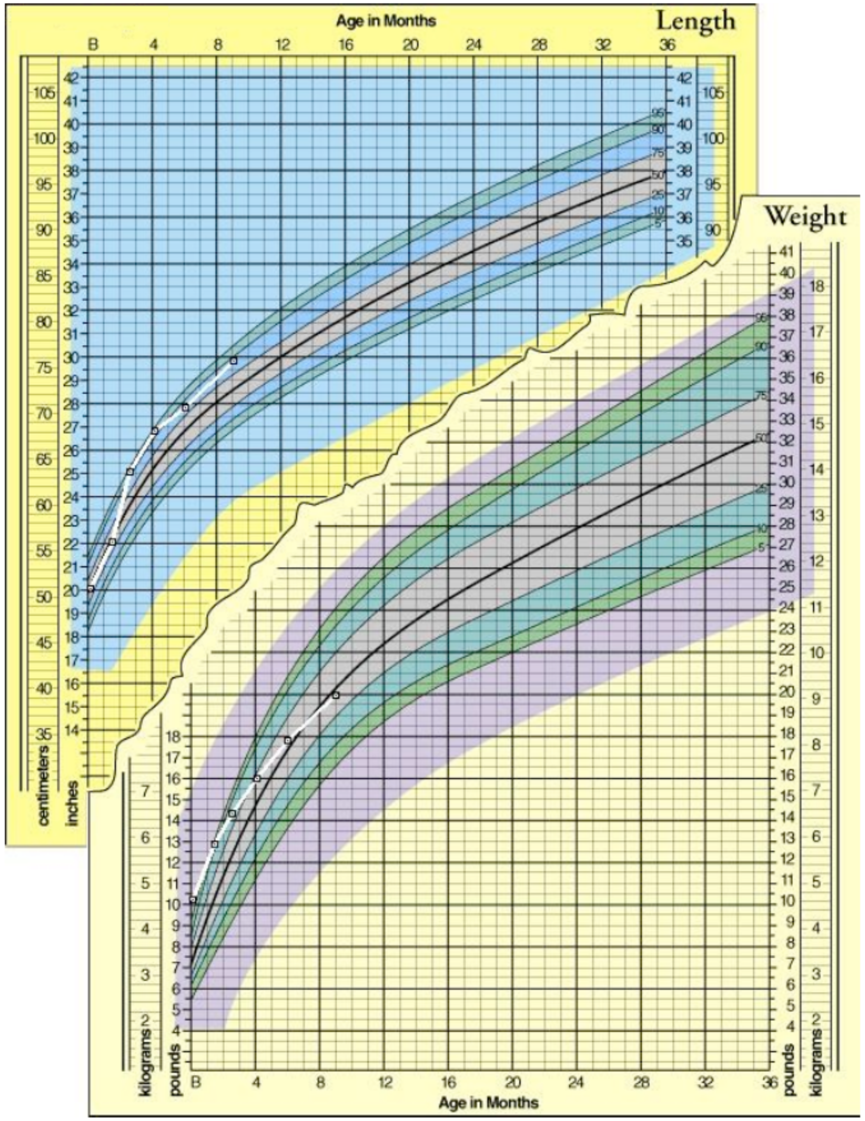
Possible extensions to functional bagplot. Consider a set of functions $f_i(x)$, $i = 1, \cdots, n$, such that

$$f_i(x) = \mu(x) + \sum_{k=1}^{n-1} z_{i,k}\varphi_k(x)$$

(i.e. principal component decomposition) where $\varphi_k(\cdot)$ represents the eigenfunctions. Rousseeuw et al., 1999 considered bivariate depth on the first two scores, $\boldsymbol{x}_i = (z_{i,1}, z_{i,2})$. See Ferraty & Vieu (2006, Nonparametric Functional Data).

Quantiles and Quantile Regressions

Quantiles are important quantities in many areas (inequalities, risk, health, sports, etc).



Quantiles of the $\mathcal{N}(0,1)$ distribution.

A First Model for Conditional Quantiles

Consider a location model, $y = \beta_0 + \boldsymbol{x}^\top \boldsymbol{\beta} + \varepsilon$ i.e.

$$\mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}] = \boldsymbol{x}^\top \boldsymbol{\beta}$$

then one can consider

$$Q(\tau|\boldsymbol{X} = \boldsymbol{x}) = \beta_0 + Q_\varepsilon(\tau) + \boldsymbol{x}^\top \boldsymbol{\beta}$$

where $Q_\varepsilon(\cdot)$ is the quantile function of the residuals.

OLS Regression, ℓ_2 norm and Expected Value

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

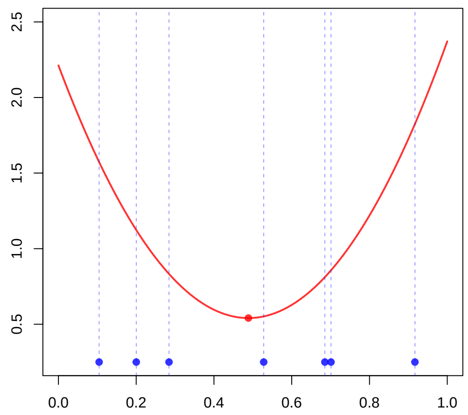
$$\mathbb{E}[Y] = \underset{m\in\mathbb{R}}{\text{argmin}} \left\{ \int \big[\underbrace{y - m}_{\varepsilon}\big]^2 dF(y) \right\} = \underset{m\in\mathbb{R}}{\text{argmin}} \left\{ \mathbb{E}\big[\|\underbrace{Y - m}_{\varepsilon}\|_{\ell_2}\big] \right\}$$

where Y is a random variable.

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

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

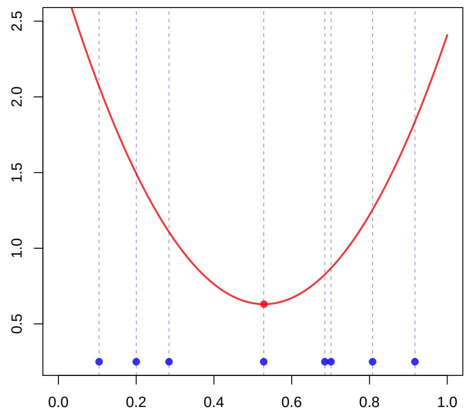
OLS Regression, ℓ_2 norm and Expected Value



Sketch of proof: (1) Let $h(x)=\displaystyle\sum_{i=1}^d (x-y_i)^2$, then

$$h'(x)=\sum_{i=1}^d 2(x-y_i)$$

and the FOC yields $x=\dfrac{1}{n}\displaystyle\sum_{i=1}^d y_i=\overline{y}$.



(2) If Y is continuous, let $h(x)=\displaystyle\int_{\mathbb{R}}(x-y)f(y)dy$ and

$$h'(x)=\frac{\partial}{\partial x}\int_{\mathbb{R}}(x-y)^2f(y)dy=\int_{\mathbb{R}}\frac{\partial}{\partial x}(x-y)^2f(y)dy$$

i.e. $x=\displaystyle\int_{\mathbb{R}}xf(y)dy=\int_{\mathbb{R}}yf(y)dy=\mathbb{E}[Y]$

Median Regression, ℓ_1 norm and Median

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

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

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

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

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

Median Regression, ℓ_1 norm and Median

Sketch of proof: (1) Let $h(x) = \displaystyle\sum_{i=1}^{d} |x - y_i|$

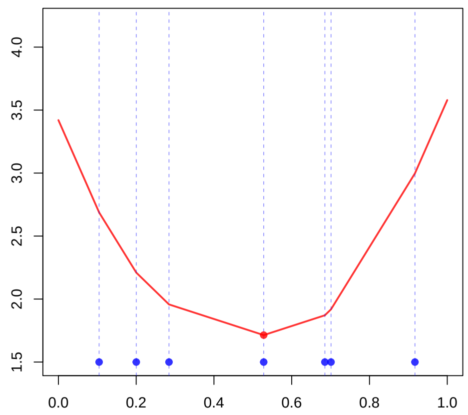
(2) If F is absolutely continuous, $dF(x) = f(x)dx$, and the median m is solution of $\displaystyle\int_{-\infty}^{m} f(x)dx = \frac{1}{2}$.

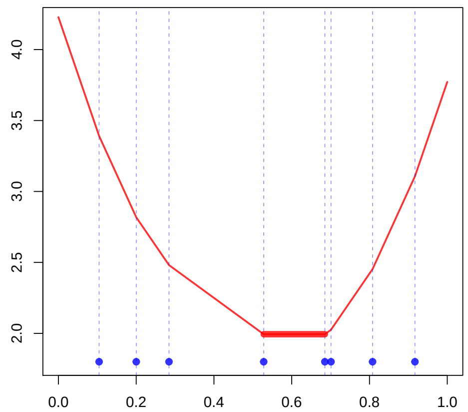
Set $h(y) = \displaystyle\int_{-\infty}^{+\infty} |x - y| f(x)dx$

$$= \int_{-\infty}^{y} (-x + y) f(x)dx + \int_{y}^{+\infty} (x - y) f(x)dx$$

Then $h'(y) = \displaystyle\int_{-\infty}^{y} f(x)dx - \int_{y}^{+\infty} f(x)dx$, and FOC yields

$$\int_{-\infty}^{y} f(x)dx = \int_{y}^{+\infty} f(x)dx = 1 - \int_{-\infty}^{y} f(x)dx = \frac{1}{2}$$





Quantiles and optimization : numerical aspects

Consider the case of the median. Consider a sample $\{y_1, \cdots, y_n\}$.

To compute the median, solve $\min_{\mu}\left\{\sum_{i=1}^n |y_i-\mu|\right\}$ which can be solved using linear programming techniques., see Dantzig (1963, Linear programming).

More precisely, this problem is equivalent to $\min_{\mu,\boldsymbol{a},\boldsymbol{b}}\left\{\sum_{i=1}^n a_i+b_i\right\}$ with $a_i, b_i \geq 0$ and $y_i-\mu=a_i-b_i$, $\forall i=1,\cdots,n$.

Consider a sample obtained from a lognormal distribution

```
n = 101
set.seed(1)
y = rlnorm(n)
median(y)
[1] 1.077415
```

Quantiles and optimization : numerical aspects

Here, one can use a standard optimization routine, `stats::optim`

```
md=Vectorize(function(m) sum(abs(y-m)))
optim(mean(y),md)
$par
[1] 1.077416
```

or a linear programing technique : use the matrix form, with $3n$ constraints, and $2n+1$ parameters, with `lpSolve::lp`

```
library(lpSolve)
A1 = cbind(diag(2*n),0)
A2 = cbind(diag(n), -diag(n), 1)
r = lp("min", c(rep(1,2*n),0),
rbind(A1, A2),c(rep("&gt;=", 2*n), rep("=", n)), c(rep(0,2*n), y))
tail(r$solution,1)
[1] 1.077415
```

Quantiles and optimization : numerical aspects

More generally, consider here some quantile,

```
tau = .3
quantile(x,tau)
      30%
0.6741586
```

The linear program is now $\min_{\mu,\boldsymbol{a},\boldsymbol{b}}\left\{\sum_{i=1}^{n}\tau a_i+(1-\tau)b_i\right\}$ with $a_i,b_i\geq 0$ and $y_i-\mu=a_i-b_i$, $\forall i=1,\cdots,n$.

```
A1 = cbind(diag(2*n),0)
A2 = cbind(diag(n), -diag(n), 1)
r = lp("min", c(rep(tau,n),rep(1-tau,n),0),
rbind(A1, A2),c(rep("&gt;=", 2*n), rep("=", n)), c(rep(0,2*n), y))
tail(r$solution,1)
[1] 0.6741586
```

Quantile regression ?

In OLS regression, we try to evaluate $\mathbb{E}[Y|\boldsymbol{X}=\boldsymbol{x}]=\int_{\mathbb{R}} ydF_{Y|\boldsymbol{X}=\boldsymbol{x}}(y)$ (using its empirical version)

In quantile regression, we try to evaluate

$$Q_u(Y|\boldsymbol{X}=\boldsymbol{x})=\inf\left\{y: F_{Y|\boldsymbol{X}=\boldsymbol{x}}(y)\geq u\right\}$$

as introduced in Newey & Powell (1987, Asymmetric Least Squares Estimation and Testing).

Li & Racine (2007, Nonparametric Econometrics: Theory and Practice) suggested

$$\widehat{Q}_u(Y|\boldsymbol{X}=\boldsymbol{x})=\inf\left\{y: \widehat{F}_{Y|\boldsymbol{X}=\boldsymbol{x}}(y)\geq u\right\}$$

where $\widehat{F}_{Y|\boldsymbol{X}=\boldsymbol{x}}(y)$ can be some smooth nonparametric estimator.

Quantiles and Expectiles

Consider the following risk functions

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

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

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

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

$$Q_Y(\tau) = \underset{m}{\operatorname{argmin}}\Big\{\mathbb{E}\big(\mathcal{R}^{\mathsf{q}}_{\tau}(Y - m)\big)\Big\}$$

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

$$E_Y(\tau) = \underset{m}{\operatorname{argmin}}\Big\{\mathbb{E}\big(\mathcal{R}^{\mathsf{e}}_{\tau}(X - m)\big)\Big\}$$

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

Quantiles and Expectiles

One can also write

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

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

Expectiles are unique, not quantiles...

Quantiles satisfy $\tau\mathbb{E}\big[\mathbf{1}_{Y-q_Y(\tau)>0}\big]=(1-\tau)\mathbb{E}\big[\mathbf{1}_{Y-q_Y(\tau)<0}\big]$

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

(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 β of

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

, and the M-estimator is the solution of

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

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

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

The continuous version is that X and Y are comonotonic if

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

One can prove that

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

Expectiles as Quantiles

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

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

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

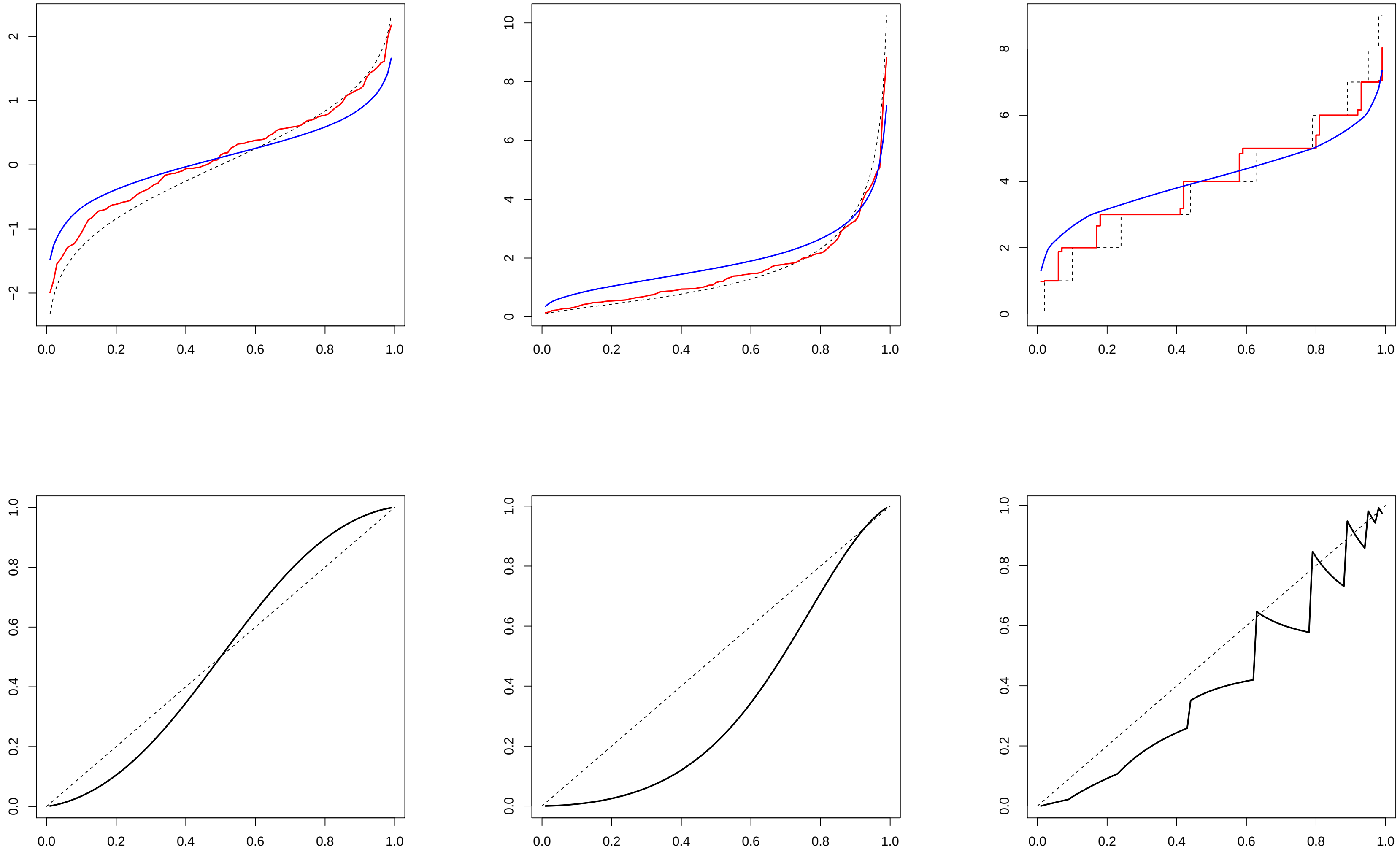
"*Expectiles have properties that are similar to quantiles*" Newey & Powell (1987, Asymmetric Least Squares Estimation and Testing). The reason is that expectiles of a distribution F are quantiles a distribution G which is related to F, see Jones (1994, Expectiles and M-quantiles are quantiles): let

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

The expectiles of F are the quantiles of G.

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

Expectiles as Quantiles



Elicitable Measures

"*elicitable*" means "being a minimizer of a suitable expected score"

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

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

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

Elicitable Measures

Remark: all functionals are not necessarily elicitable, see Osband (1985, Providing incentives for better cost forecasting)

The variance is not elicitable

The elicitability property implies a property which is known as convexity of the level sets with respect to mixtures (also called Betweenness property) : if two lotteries F, and G are equivalent, then any mixture of the two lotteries is also equivalent with F and G.

Empirical Quantiles

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

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

Then as $n \to \infty$

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

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

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

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

Empirical Expectiles

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

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

Then as $n \to \infty$

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

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

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

so that μ_τ is solution of $\mathbb{E}\big[\mathcal{I}(\mu_\tau, Y)\big] = 0$. 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, $\min\left\{\sum_{i=1}^{n}\mathcal{R}_\tau^{\mathsf{q}}(y_i-\boldsymbol{x}_i^\mathsf{T}\beta)\right\}$

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

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

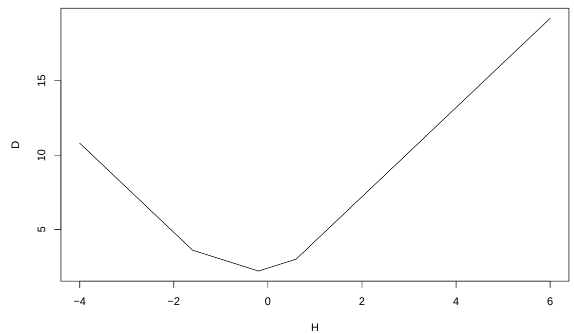
From line blue, a shift up decrease the sum by ϵ

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 piecwise linear in ϵ)



Geometric Properties of the Quantile Regression

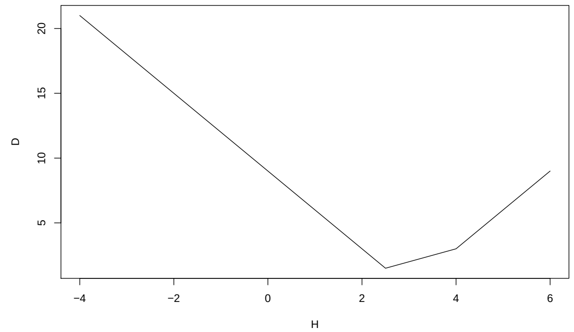
Consider now rotations of the line around the support point

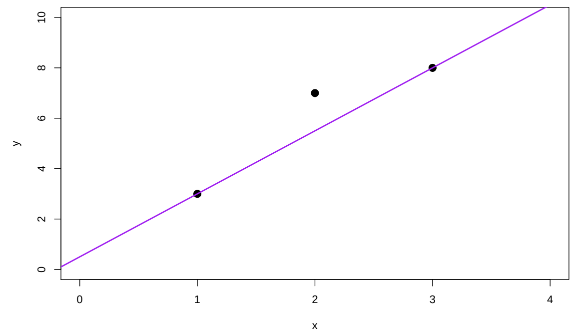
If we rotate up, we increase the sum of absolute difference (large impact on the point on the right)

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.

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





Numerical Aspects

To illustrate numerical computations, use

```
base=read.table("http://freakonometrics.free.fr/rent98_00.txt",header
    =TRUE)
```

The linear program for the quantile regression is now

$$\min_{\beta_0^\tau,\beta_1^\tau,\boldsymbol{a},\boldsymbol{b}}\left\{\sum_{i=1}^n \tau a_i+(1-\tau)b_i\right\}$$

with $a_i,b_i\geq 0$ and $y_i-[\beta_0^\tau+\beta_1^\tau x_i]=a_i-b_i$, $\forall i=1,\cdots,n$.

```
require(lpSolve)
tau = .3
n=nrow(base)
X = cbind( 1, base$area)
y = base$rent_euro
A1 = cbind(diag(2*n), 0,0)
A2 = cbind(diag(n), -diag(n), X)
```

Numerical Aspects

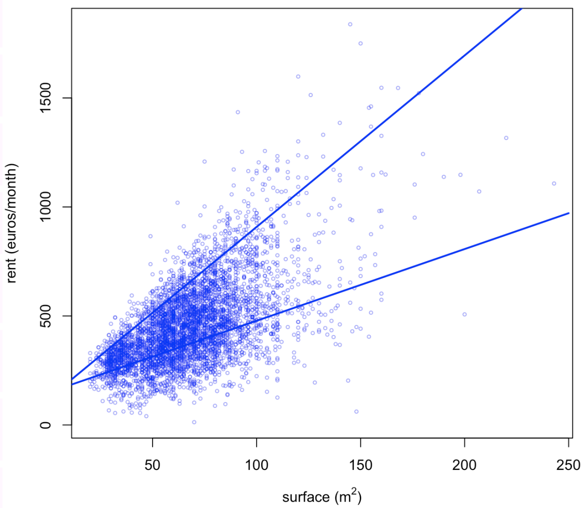
```
r = lp("min",
        c(rep(tau,n), rep(1-tau,n),0,0), rbind(A1, A2),
        c(rep("&gt;=", 2*n), rep("=", n)), c(rep(0,2*n), y))
tail(r$solution,2)
[1] 148.946864   3.289674
```

see `quantreg::rq()` function

```
library(quantreg)
rq(rent_euro~area, tau=tau, data=base)
Coefficients:
(Intercept)        area
 148.946864    3.289674
```

Numerical Aspects

```
plot(base$area,base$rent_euro)
sf=0:250
yr=r$solution[2*n+1]+r$solution[2*n+2]*sf
lines(sf,yr,lwd=2,col="blue")
tau = .9
r = lp("min",c(rep(tau,n), rep(1-tau,n),0,0),
    rbind(A1, A2),c(rep(">=", 2*n), rep("=", n)
    ), c(rep(0,2*n), y))
yr=r$solution[2*n+1]+r$solution[2*n+2]*sf
lines(sf,yr,lwd=2,col="blue")
```



Numerical Aspects

For multiple regression, we should consider some trick (R function assumes all variables are nonnegative)

```
 tau = 0.3
n = nrow(base)
X = cbind(1, base$area, base$yearc)
y = base$rent_euro
r = lp("min",
c(rep(tau, n), rep(1 - tau, n), rep(0, 2 * 3)),
cbind(diag(n), -diag(n), X, -X),
rep("=", n),
y)
beta = tail(r$solution, 6)
beta = beta[1:3] - beta[3 + 1:3]
beta
[1]  -5542.503252     3.978135     2.887234
```

Numerical Aspects

which is consistant with the output from `quantreg::rq`

```
library(quantreg)
rq(rent_euro~area+yearc, tau=tau, data=base)
Coefficients:
 (Intercept)          area        yearc
-5542.503252      3.978135     2.887234
```

Distributional Aspects

OLS are equivalent to MLE when $Y - m(\boldsymbol{x}) \sim \mathcal{N}(0, \sigma^2)$, 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 $Y - m(\boldsymbol{x})$ has an asymmetric Laplace distribution

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

Quantile Regression and Iterative Least Squares

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

at stage k :

- let $\varepsilon_i^{(k)} = y_i - \boldsymbol{x}_i^\mathsf{T}\boldsymbol{\beta}^{(k-1)}$
- define weights $\omega_i^{(k)} = \mathcal{R}'_\tau(\varepsilon_i^{(k)})$
- compute weighted least square to estimate $\boldsymbol{\beta}^{(k)}$

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

Optimization Algorithm

Primal problem is

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

and the dual version is

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

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.

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 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 $k = \dim(\boldsymbol{\beta})$ (it is $(n+k)k^2$ for OLS, see wikipedia).

Quantile Regression Estimators

OLS estimator $\widehat{\boldsymbol{\beta}}^{\mathsf{ols}}$ is solution of

$$\widehat{\boldsymbol{\beta}}^{\mathsf{ols}} = \text{argmin}\left\{\mathbb{E}\left[\left(\mathbb{E}[Y|\boldsymbol{X}=\boldsymbol{x}] - \boldsymbol{x}^{\mathsf{T}}\boldsymbol{\beta}\right)^2\right]\right\}$$

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

$$\widehat{\boldsymbol{\beta}}_\tau = \text{argmin}\left\{\mathbb{E}\left[\omega_\tau(\boldsymbol{\beta})\left(Q_\tau[Y|\boldsymbol{X}=\boldsymbol{x}] - \boldsymbol{x}^{\mathsf{T}}\boldsymbol{\beta}\right)^2\right]\right\}$$

(under weak conditions) where

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

$\widehat{\boldsymbol{\beta}}_\tau$ is the best weighted mean square approximation of the tru quantile function, where the weights depend on an average of the conditional density of Y over $\boldsymbol{x}^{\mathsf{T}}\boldsymbol{\beta}$ 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 $(Y_i, \boldsymbol{X}_i)$ must (conditionnaly) i.id.
- regressors must have a bounded second moment, $\mathbb{E}\big[\|\boldsymbol{X}_i\|^2\big] < \infty$
- error terms ε are continuously distributed given $\boldsymbol{X}_i$, centered in the sense that their median should be 0,

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

- "local identification" property : $\big[f_\varepsilon(0)\boldsymbol{X}\boldsymbol{X}^\mathsf{T}\big]$ is positive definite

Quantile Regression Estimators

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

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{\beta}_\tau)\overset{\mathcal{L}}{\rightarrow}\mathcal{N}(0,\tau(1-\tau)D_\tau^{-1}\Omega_x D_\tau^{-1}),$$

where

$$D_\tau=\mathbb{E}\big[f_\varepsilon(0)\boldsymbol{X}\boldsymbol{X}^\mathsf{T}\big]\text{ and }\Omega_x=\mathbb{E}\big[\boldsymbol{X}^\mathsf{T}\boldsymbol{X}\big].$$

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

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

where $\widehat{f}_\varepsilon(0)$ is estimated using (e.g.) an histogram, as suggested in Powell (1991, Estimation of monotonic regression models under quantile restrictions), since

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

Quantile Regression Estimators

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

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

There is an asymptotic first order condition,

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

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

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

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

- generate a sample $(y_i^{(b)}, \boldsymbol{x}_i^{(b)})$ from $(y_i, \boldsymbol{x}_i)$
- estimate $\boldsymbol{\beta}_\tau^{(b)}$ by
$$\widehat{\boldsymbol{\beta}}_\tau^{(b)} = \text{argmin}\left\{\mathcal{R}_\tau^{\mathsf{q}}\left(y_i^{(b)} - \boldsymbol{x}_i^{(b)\mathsf{T}}\boldsymbol{\beta}\right)\right\}$$
- set $\widehat{\text{Var}}^\star\big[\widehat{\boldsymbol{\beta}}_\tau\big] = \dfrac{1}{B}\displaystyle\sum_{b=1}^{B}\big(\widehat{\boldsymbol{\beta}}_\tau^{(b)} - \widehat{\boldsymbol{\beta}}_\tau\big)^2$

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

Quantile Regression Estimators

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

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

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 $\tau \in [0,1]$

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

Equivariance to reparameterization of design

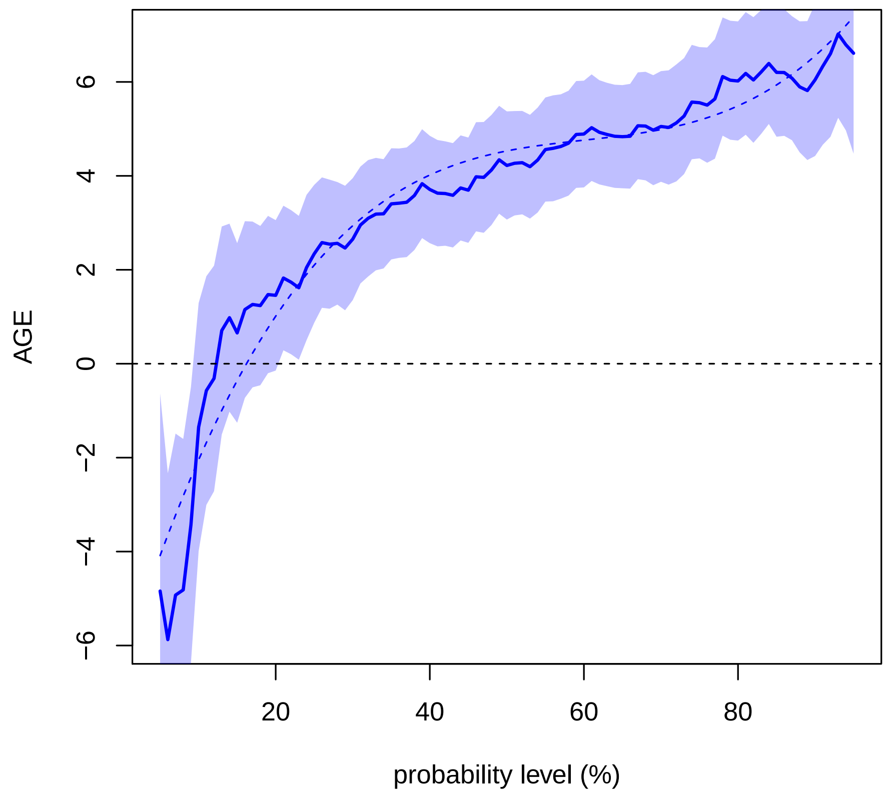
Let $\boldsymbol{A}$ be any $p \times p$ nonsingular matrix and $\tau \in [0,1]$

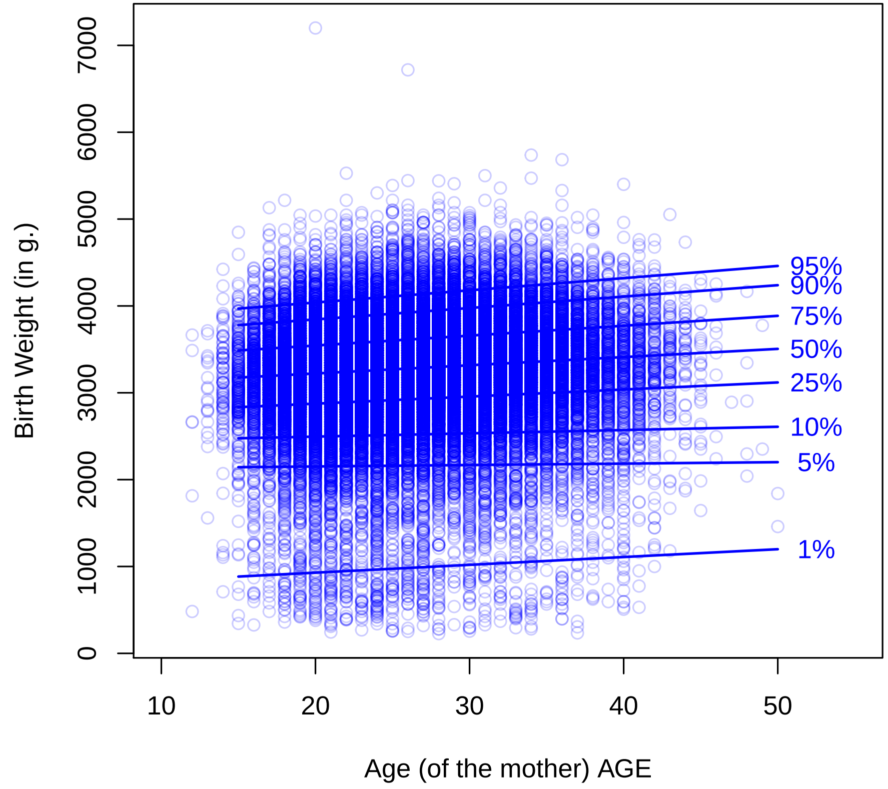
$$\hat{\boldsymbol{\beta}}_\tau(Y, \boldsymbol{X}\boldsymbol{A}) = \boldsymbol{A}^{-1}\hat{\boldsymbol{\beta}}_\tau(Y, \boldsymbol{X})$$

Visualization, $\tau \mapsto \widehat{\beta}_\tau$

See Abreveya (2001, The effects of demographics and maternal behavior...)

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



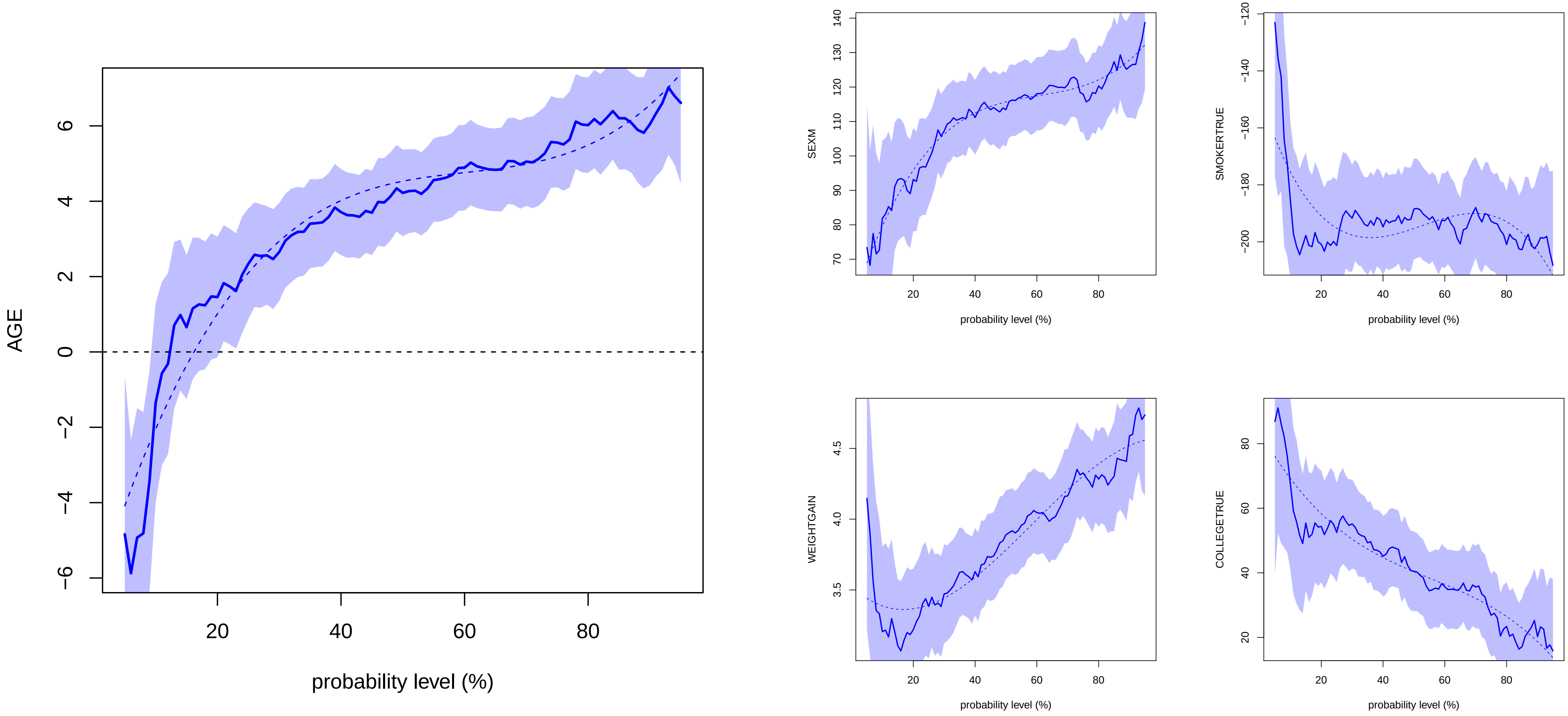


Visualization, $\tau \mapsto \widehat{\beta}_\tau$

```
> base=read.table("http://freakonometrics.free.fr/natality2005.txt",
    header=TRUE,sep=";")
> u=seq(.05,.95,by=.01)
> library(quantreg)
>  coefstd=function(u) summary(rq(WEIGHT~SEX+SMOKER+WEIGHTGAIN+
    BIRTHRECORD+AGE+ BLACKM+ BLACKF+COLLEGE,data=sbase,tau=u))$
    coefficients[,2]
> coefest=function(u) summary(rq(WEIGHT~SEX+SMOKER+WEIGHTGAIN+
    BIRTHRECORD+AGE+ BLACKM+ BLACKF+COLLEGE,data=sbase,tau=u))$
    coefficients[,1]
CS=Vectorize(coefstd)(u)
CE=Vectorize(coefest)(u)
```

Visualization, $\tau \mapsto \widehat{\beta}_\tau$

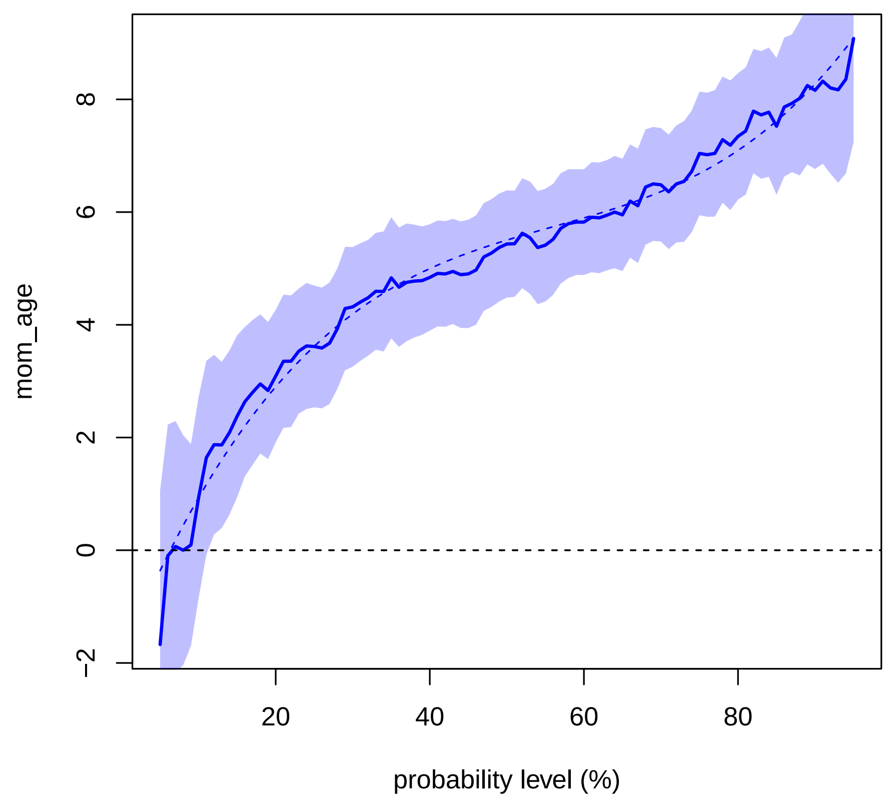
See Abreveya (2001, The effects of demographics and maternal behavior on the distribution of birth outcomes)

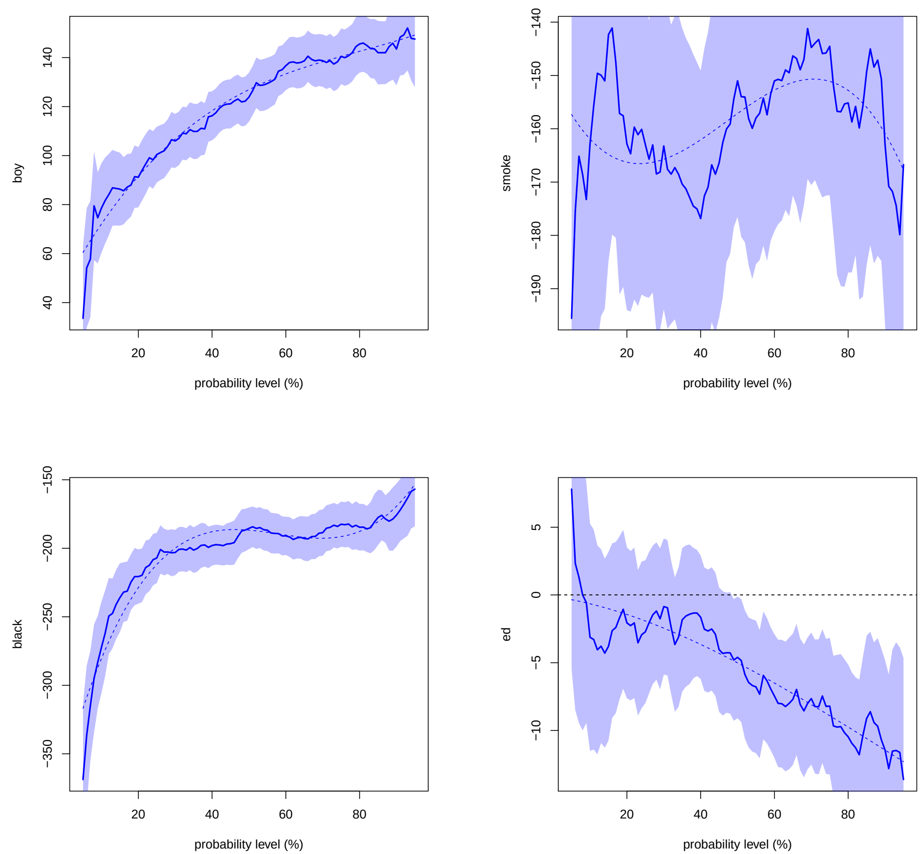


Visualization, $\tau \mapsto \widehat{\beta}_\tau$

See Abreveya (2001, The effects of demographics and maternal behavior...)

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

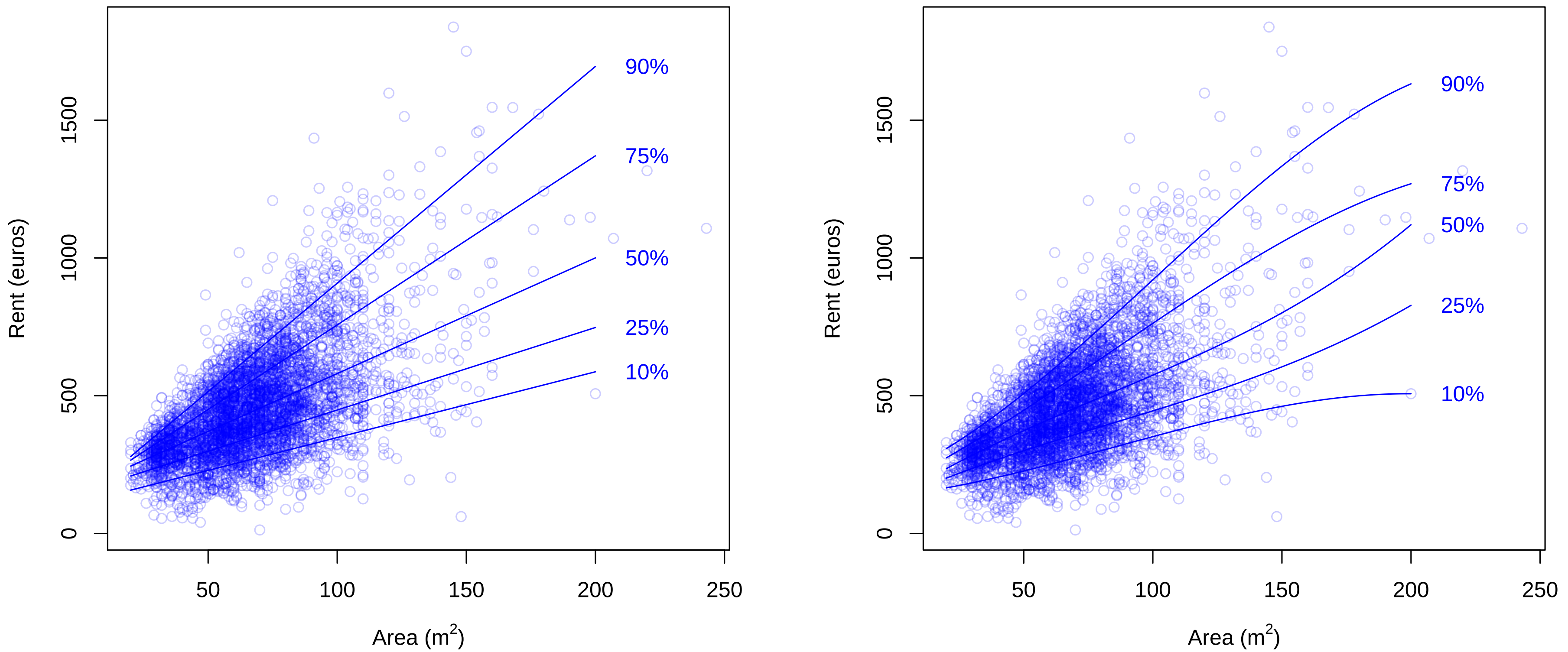




Quantile Regression, with Non-Linear Effects

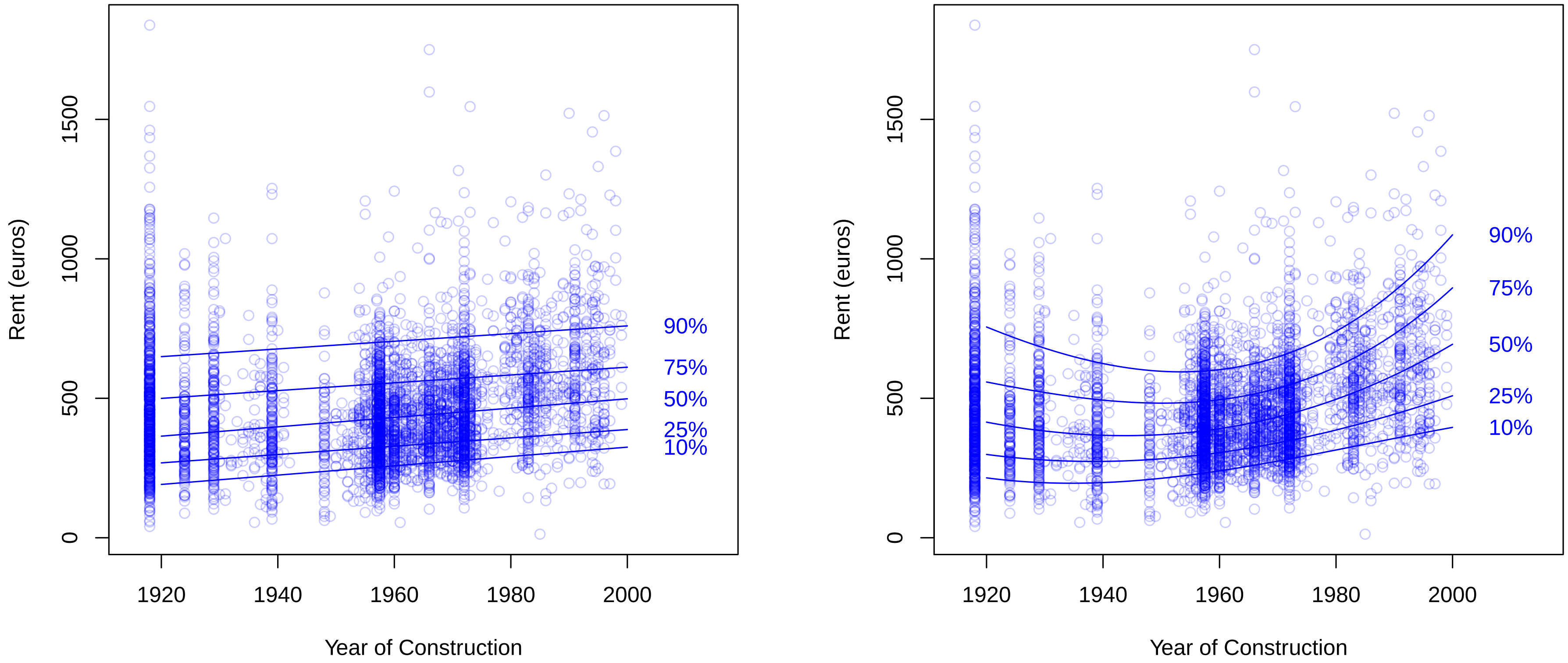
Rents in Munich, as a function of the area, from Fahrmeir *et al.* (2013, Regression: Models, Methods and Applications)

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



Quantile Regression, with Non-Linear Effects

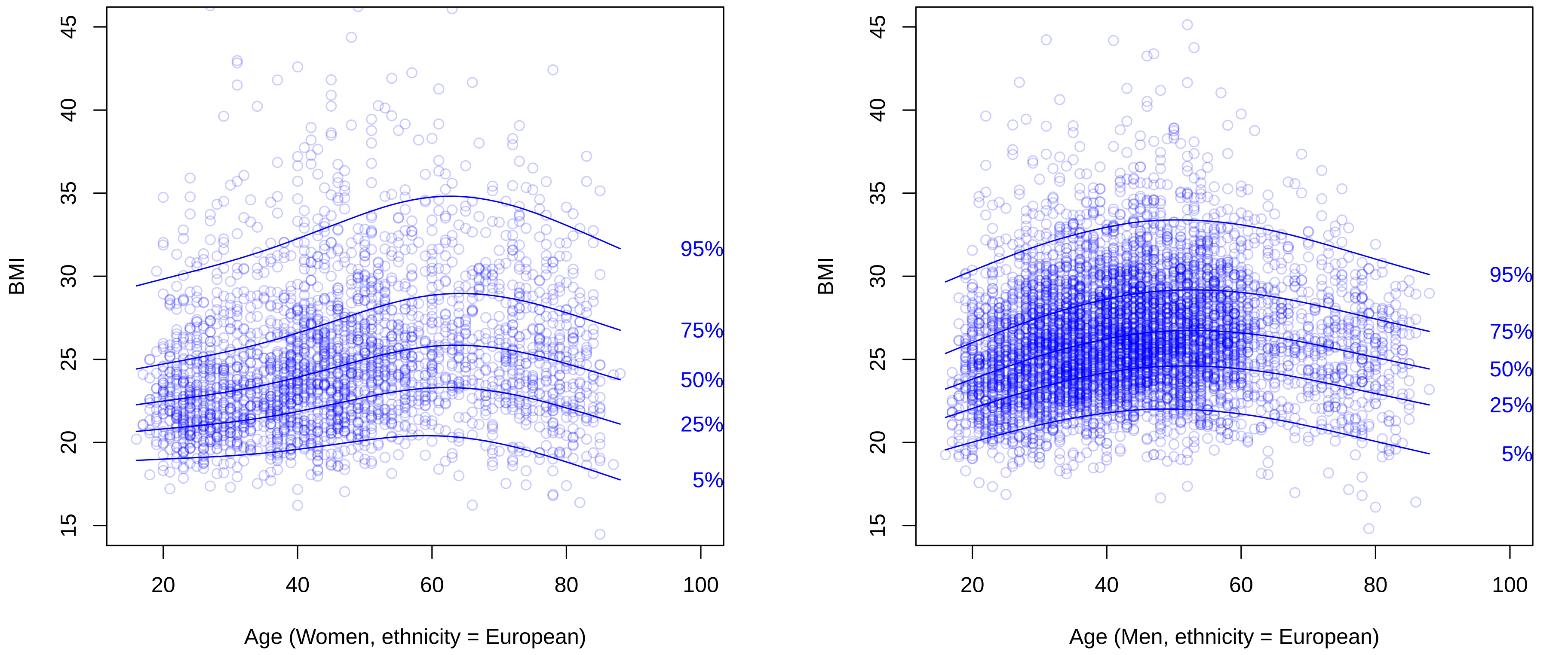
Rents in Munich, as a function of the year of construction, from Fahrmeir *et al.* (2013, Regression: Models, Methods and Applications)



Quantile Regression, with Non-Linear Effects

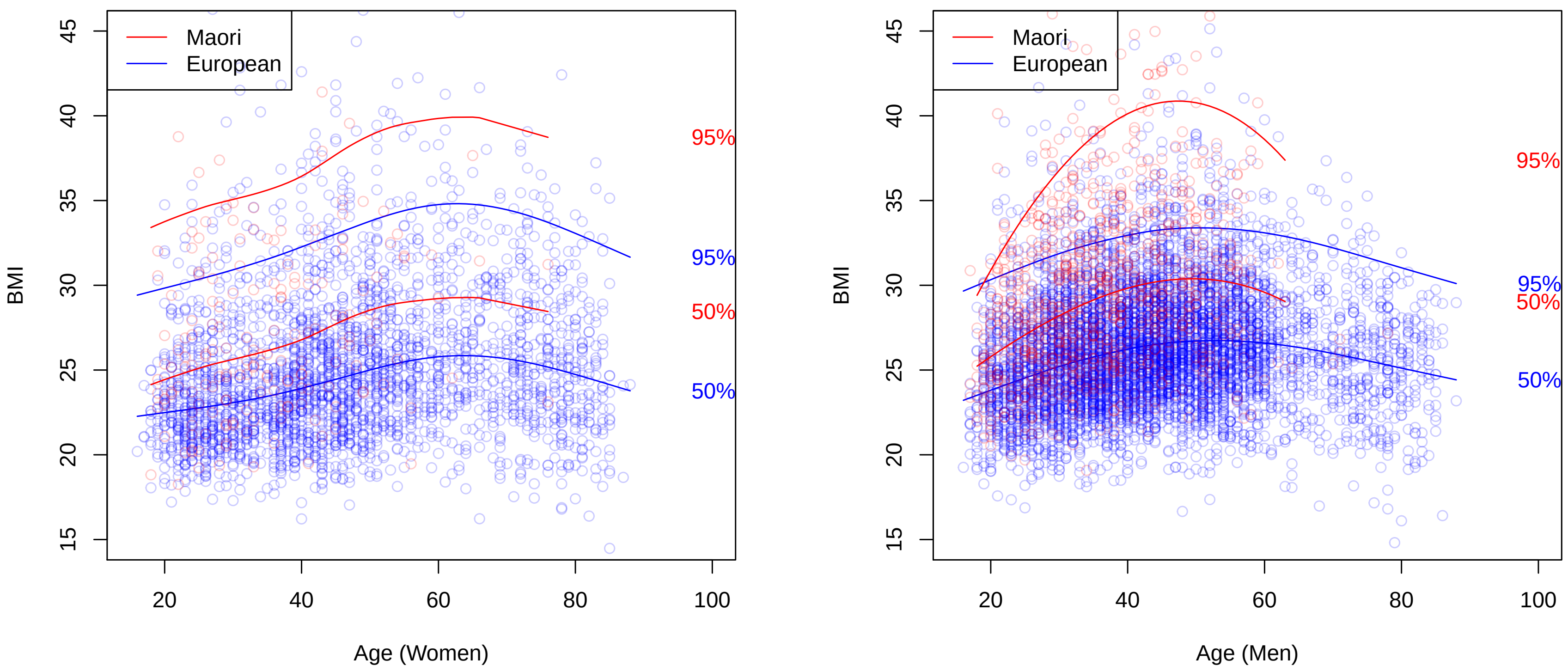
BMI as a function of the age, in New-Zealand, from Yee (2015, Vector Generalized Linear and Additive Models), for Women and Men

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


BMI
Age (Women, ethnicity = European)
95%
75%
50%
25%
5%
BMI
Age (Men, ethnicity = European)
95%
75%
50%
25%
5%

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.

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

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, $y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$. Let

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

One might consider $\omega_\alpha = 1 + \dfrac{z_\alpha}{\varphi(z_\alpha) + (1-\alpha) z_\alpha}$ 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

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

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

Non-Crossing Solutions

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

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

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

$$\boldsymbol{x}_i^\mathsf{T}\widehat{\boldsymbol{\beta}}_{\tau_j} \geq \boldsymbol{x}_i^\mathsf{T}\widehat{\boldsymbol{\beta}}_{\tau_{j-1}} \quad \forall i \in \{1, \cdots, n\}, j \in \{2, \cdots, q\}.$$

Quantile Regression for Time Series

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

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

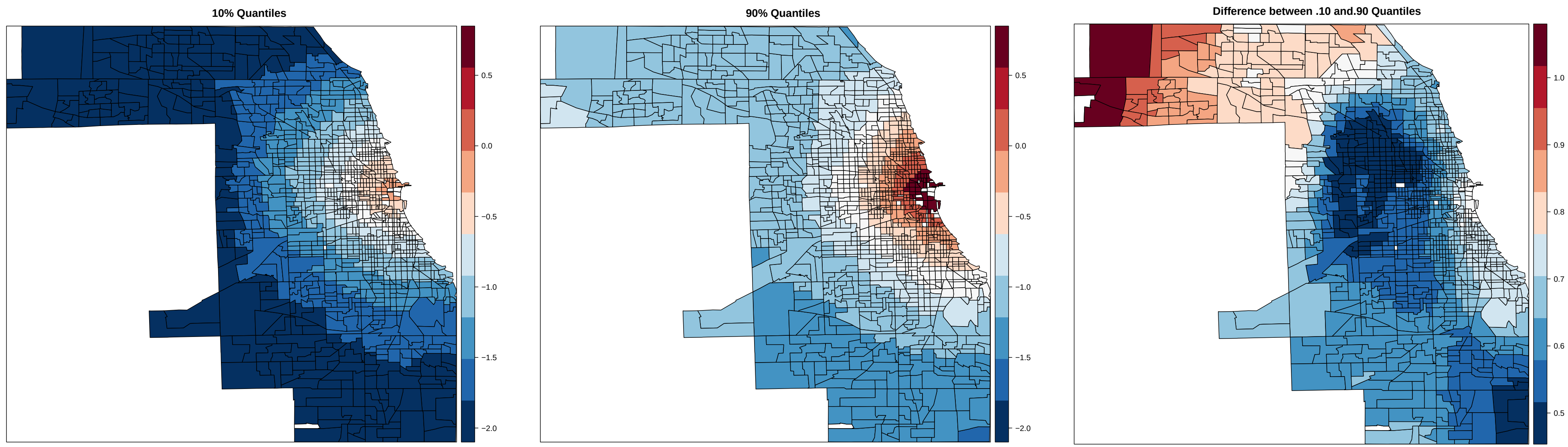
The quantile function conditional on the past - $\mathcal{F}_{t-1} = \underline{Y}_{t-1}$ - is

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

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)

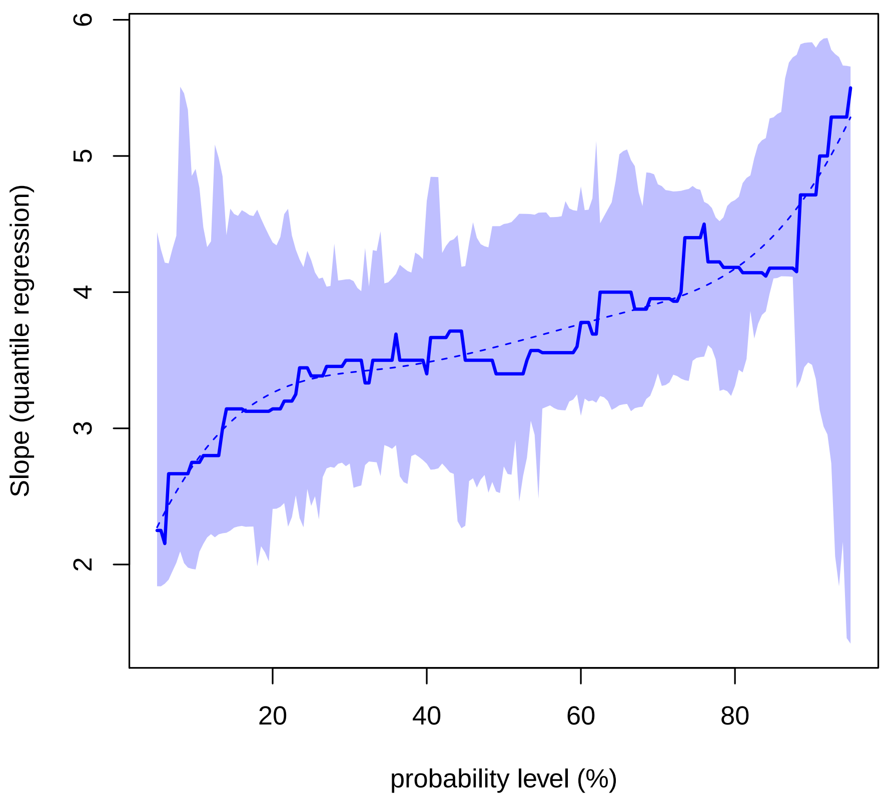
Quantile Regression for Spatial Data

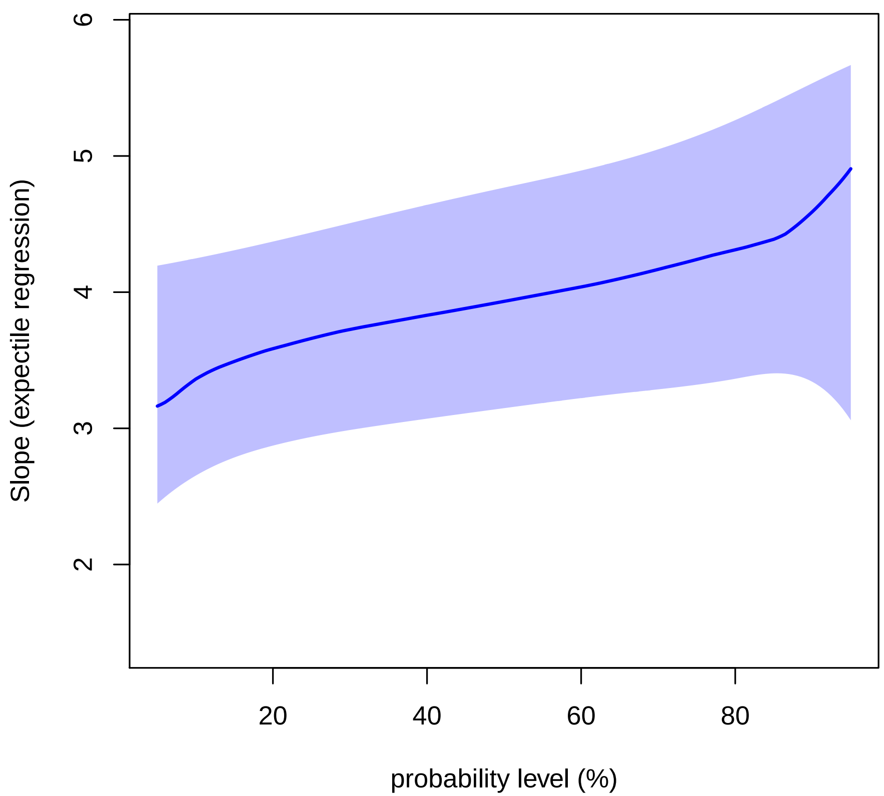
```
> library(McSpatial)
> data(cookdata)
> fit <- qregcpar(LNFAR~DCBD, nonpar=~LATITUDE+LONGITUDE, taumat=c
    (.10,.90), kern="bisq", window=.30, distance="LATLONG", data=
    cookdata)
```



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

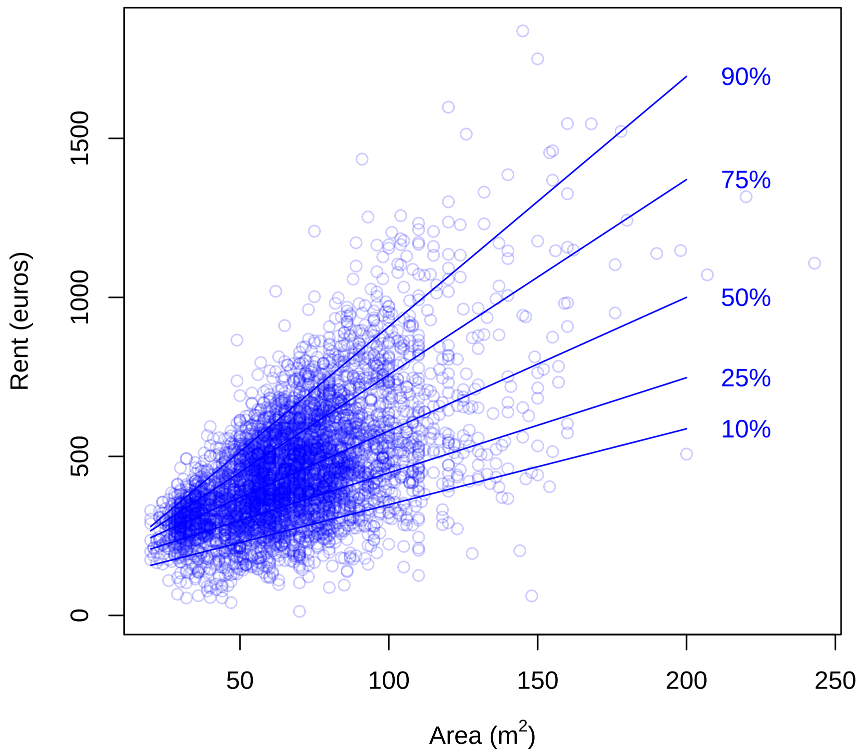
"*this estimator can be interpreted as a maximum likelihood estimator when the disturbances arise from a normal distribution with unequal weight placed on positive and negative disturbances*" Aigner, Amemiya & Poirier (1976, Formulation and Estimation of Stochastic Frontier Production Function Models).

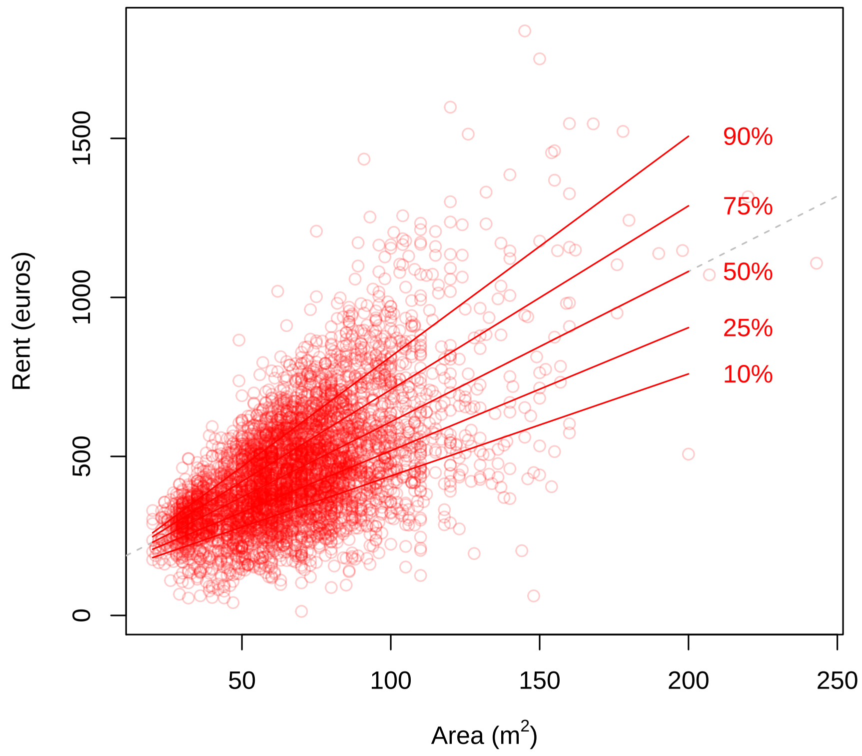
See Holzmann & Klar (2016, Expectile Asymptotics) for statistical properties.

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



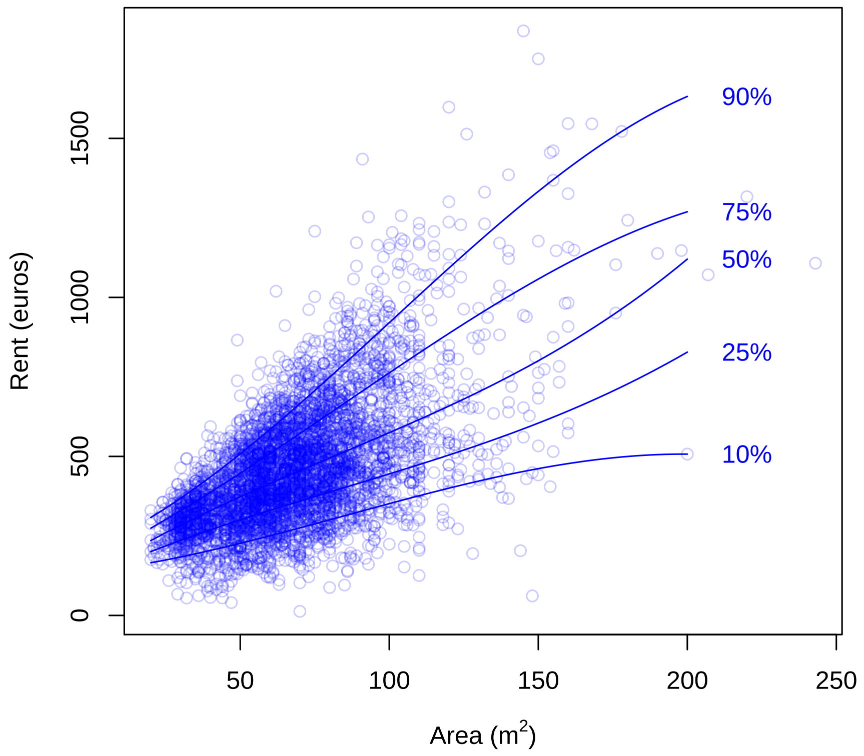


Quantile Regressions

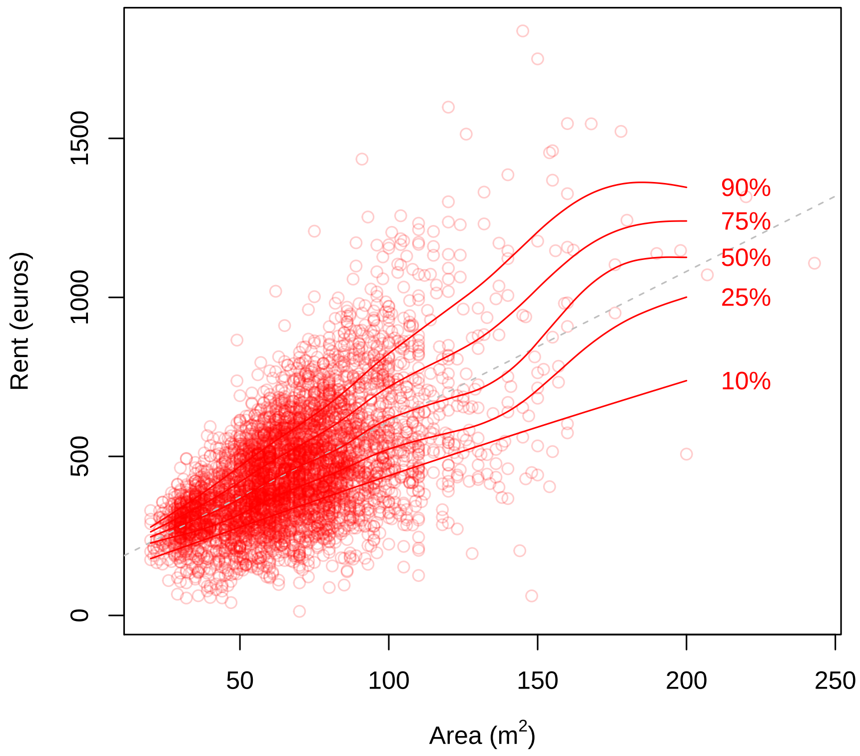
Expectile Regressions

Expectile Regression, with Non-Linear Effects

See Zhang (1994, Nonparametric regression expectiles)



Quantile Regressions



Expectile Regressions

Expectile Regression, with Linear Effects

```
> library(expectreg)
> coefstd=function(u) summary(expectreg.ls(WEIGHT~SEX+SMOKER+
    WEIGHTGAIN+BIRTHRECORD+AGE+ BLACKM+ BLACKF+COLLEGE,data=sbase,
    expectiles=u,ci = TRUE))[,2]
> coefest=function(u) summary(expectreg.ls(WEIGHT~SEX+SMOKER+
    WEIGHTGAIN+BIRTHRECORD+AGE+ BLACKM+ BLACKF+COLLEGE,data=sbase,
    expectiles=u,ci = TRUE))[,1]
> CS=Vectorize(coefstd)(u)
> CE=Vectorize(coefest)(u)
```