# Additive quantile regression with qgam

Matteo Fasiolo (University of Bristol, UK)

Joint work with:

Simon N. Wood (University of Bristol, UK)
Yannig Goude (EDF R&D)
Raphael Nedellec (EDF R&D)

*matteo.fasiolo@bristol.ac.uk*

Oct 19, 2017

# Structure of the seminar

**Structure:**

1. Distributional vs quantile regression
2. Additive quantile regression
3. The `qgam` R package

# Structure of the seminar

**Structure:**

1. **Distributional vs quantile regression**
2. Additive quantile regression
3. The qgam R package

# Distributional vs quantile regression

In **distributional regression** we want a good model for $p(y|\mathbf{x})$.

We indicate it with $p_m\{y|\theta_1(\mathbf{x}), \dots, \theta_q(\mathbf{x})\}$, where $\theta_1(\mathbf{x}), \dots, \theta_q(\mathbf{x})$ are model parameters.

For instance, if we use a Gaussian model, we can make its mean and/or variance depend on the covariates, that is

$$y|\mathbf{x} \sim N\big\{y|\mu = \theta_1(\mathbf{x}), \sigma^2 = \theta_2(\mathbf{x})\big\},$$

where $\mu = \mathbb{E}(y|\mathbf{x})$ and $\sigma^2 = \mathrm{Var}(y|\mathbf{x})$.

# Distributional vs quantile regression

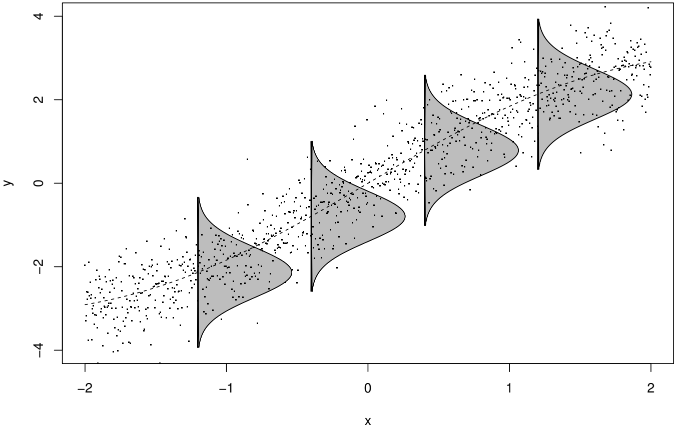



Figure: We are using of Gaussian model with variable mean.
In mgcv: `gam(y~s(x), family=gaussian)`.

# Distributional vs quantile regression

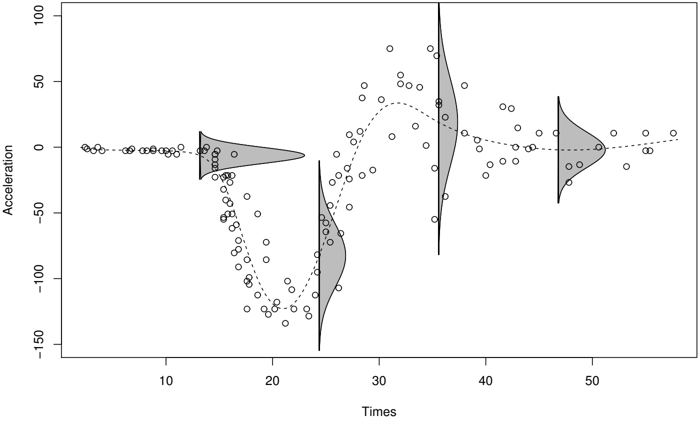
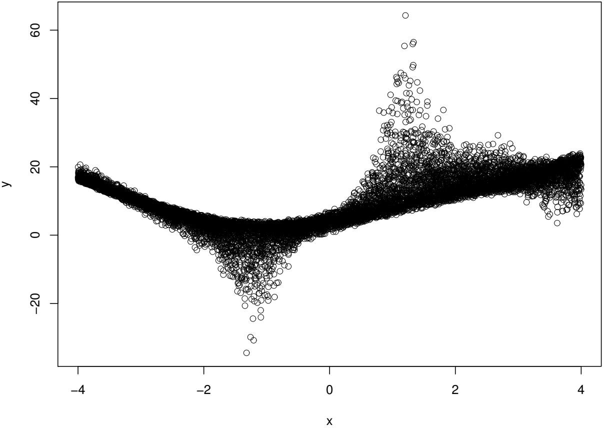



Figure: We are using of Gaussian model with variable mean and variance.
In mgcv: `gam(list(y~s(x), ~s(x)), family=gaulss)`.

# Distributional vs quantile regression

There are lots of models for $p(y|\mathbf{x})$ that can be used for distributional regression: binomial, gamma, Poisson, Tweedie, etc...

But sometimes it is difficult to find a good model.

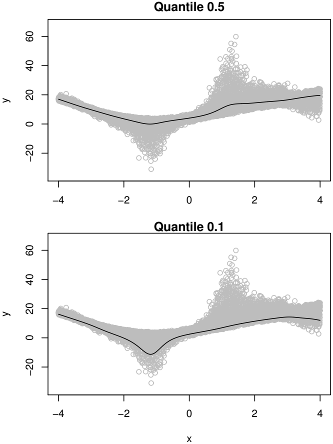


# Distributional vs quantile regression

In QR we do not have a model for $p(y|\mathbf{x})$, but we model quantiles directly.

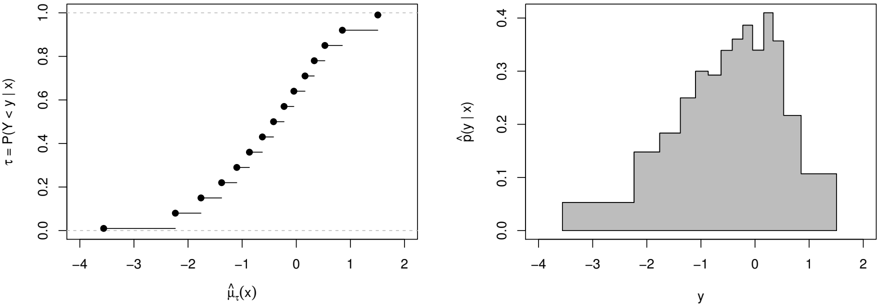
Let $F(y|\mathbf{x})$ be Prob$(Y \leq y|\mathbf{x})$. The $\tau$-th ($\tau \in (0,1)$) quantile is

$$\mu_\tau(\mathbf{x}) = F^{-1}(\tau|\mathbf{x}).$$



# Distributional vs quantile regression

Obviously, given a set of fitted quantiles, we can approximate $p(y|\mathbf{x})$.



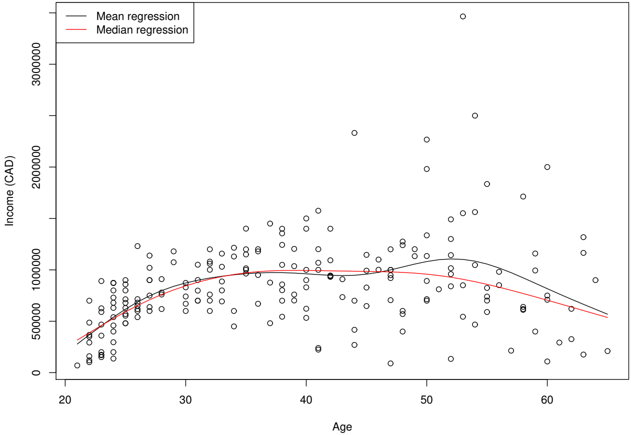

So we can use QR when we can't model $p(y|\mathbf{x})$ directly.

But we could be genuinely interested in estimating a quantile $\mu_\tau(\mathbf{x})$.

# Distributional vs quantile regression

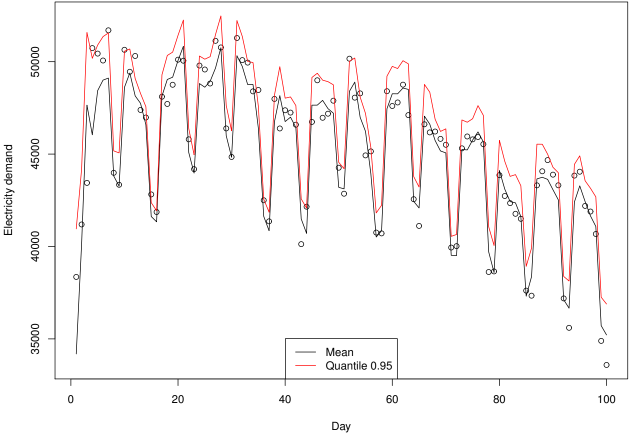
We might want to model median, rather than mean, income because:

a) it is more robust to outliers then mean income;
b) it is a better indicator of how the "average" person is doing.



# Distributional vs quantile regression

For electricity producers and/or distributors the top electricity demand on a given day might be more important that mean demand.



# Structure of the seminar

**Structure:**

1. Distributional vs quantile regression
2. **Additive quantile regression**
3. The qgam R package
4. Limitations and future developments

# Additive quantile regression

We are in an additive framework

$$\mu_\tau(\mathbf{x}) = \sum_{j=1}^{m} f_j(\mathbf{x}),$$

where the effects $f_j(\mathbf{x})$ can be parametric, smooth or random effects.

Given $\tau$, how is $\mu_\tau(\mathbf{x})$ estimated?

Key fact is that $\mu_\tau(\mathbf{x})$ is minimizer of

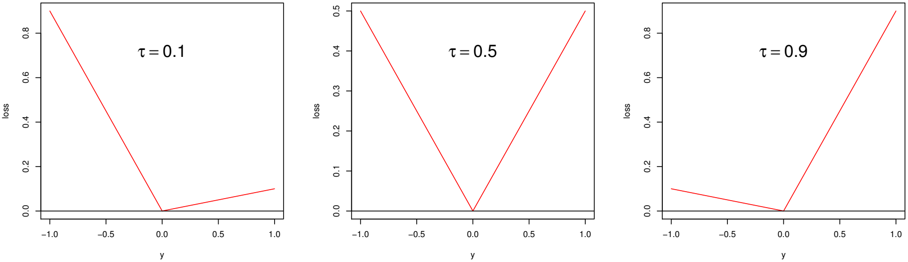
$$L_\tau\{\mu(\mathbf{x})\} = \mathbb{E}\{\rho_\tau\{y - \mu(\mathbf{x})\}|\mathbf{x}\},$$

where

$$\rho_\tau(z) = (\tau - 1)z\mathbb{1}(z < 0) + \tau z\mathbb{1}(z \geq 0),$$

is the so-call "check" function or "pinball" loss.

# Additive quantile regression




Given $n$ observations, $\{y_1, \mathbf{x}_1\}, \dots, \{y_n, \mathbf{x}_n\}$, the total loss is

$$\hat{\mu}_\tau(\mathbf{x}) = \underset{\mu}{argmin} \sum_{i=1}^{n} \rho_\tau[y_i - \mu(\mathbf{x}_i)],$$

hence we could estimate $\mu_\tau(\mathbf{x})$ by minimizing

$$PL\{\mu(\mathbf{x})\} = L\{\mu(\mathbf{x})\} + \sum_{j=1}^{m} \gamma_j \int f_j''(x_j)^2 dx_j.$$

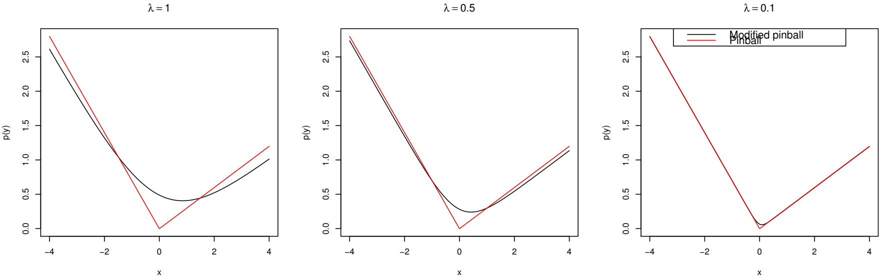
but $L\{\mu(\mathbf{x})\}$ is not differentiable!

# Additive quantile regression

We use a modified loss based on the Extended log-F (ELF) density:

$$\tilde{\rho}(y|\mu,\sigma,\tau,\lambda) \propto -(1-\tau)\frac{y-\mu}{\sigma} + \lambda \log\left[1 + e^{\frac{y-\mu}{\lambda\sigma}}\right],$$

This is smooth and convex and, as $\lambda \to 0$, we have recover pinball loss.




NB in `qgam`, $\lambda$ reparametrized as `err` $\in (0,1)$ ($\downarrow$ `err` implies $\downarrow \lambda$).

# Additive quantile regression

For fixed $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_m\}$ we can estimate $\mu_\tau(\mathbf{x}) = \underset{\mu}{argmin}\, PL\{\mu(\mathbf{x})\}$.

But how to select $\boldsymbol{\gamma}$ in the first place?

We minimize the marginal loss

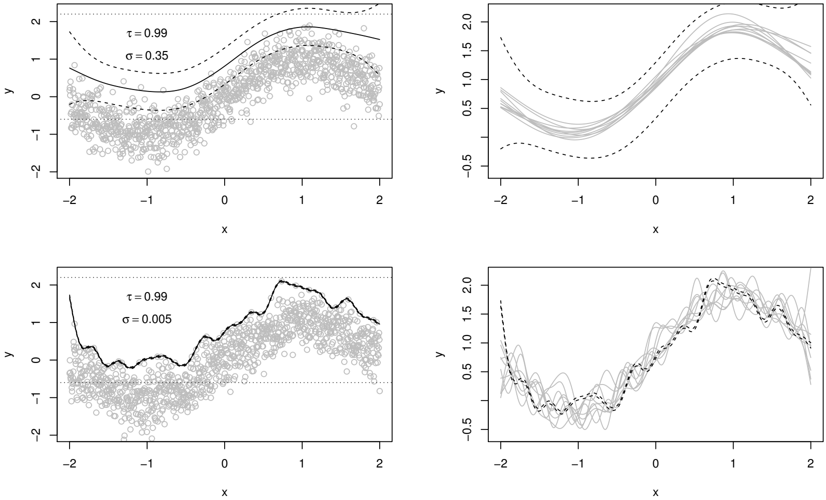
$$G(\boldsymbol{\gamma}, \sigma) = -\int \underbrace{\exp\left[-\frac{1}{\sigma} L\{\mu(\mathbf{x})\}\right]}_{\text{fit the data well}} \overbrace{p(\boldsymbol{\beta}|\boldsymbol{\gamma})}^{\text{keep fit simple}} \, d\boldsymbol{\beta}.$$

where $p(\boldsymbol{\beta}|\boldsymbol{\gamma})$ is related to the penalties $\sum_{j=1}^m \gamma_j \int f_j''(x_j)^2 dx_j$.

Here $\sigma$ is the reciprocal of the "learning rate" and it needs to be selected as well...

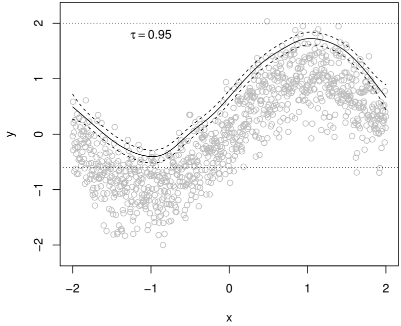
# Additive quantile regression

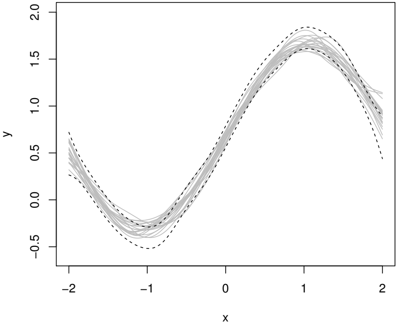
If $\uparrow \sigma$ then $\uparrow \hat{\gamma}$ (smoother fit) and wider conf. int. for $\mu_\tau(\mathbf{x})$.
If $\downarrow \sigma$ then $\downarrow \hat{\gamma}$ (more wiggly fit) and narrower conf. int. for $\mu_\tau(\mathbf{x})$.



# Additive quantile regression

We choose $\sigma$ so that confidence intervals for $\mu_\tau(\mathbf{x})$ are well calibrated.




In practice we use bootstrapping to calibrate $\sigma$.

In `qgam` this is implemented by `tuneLearn()` and `tuneLearnFast()`.

# Additive quantile regression

To recap, we have three nested iterations:

1. Calibration to estimate reciprocal of learning rate, $\sigma$.
2. Minimize ML to select smoothing parameters $\gamma$ (**fixed $\boldsymbol{\sigma}$**):

$$G(\gamma, \sigma) = -\int \exp\big[-\sigma^{-1} L\{\mu(\mathbf{x})\}\big] p(\boldsymbol{\beta}|\gamma) d\boldsymbol{\beta}.$$

3. Minimize PL to estimate regression coefficients $\boldsymbol{\beta}$ (**fixed $\boldsymbol{\sigma}$ and $\gamma$**):

$$PL\{\mu(\mathbf{x})\} = L\{\mu(\mathbf{x})\} + \sum_{j=1}^{m} \gamma_j \int f_j''(x_j)^2 dx_j.$$

For more methodology, see Fasiolo et al. (2017) on arXiv.

# Structure of the seminar

**Structure:**

1. Distributional vs quantile regression
2. Additive quantile regression
3. **The `qgam` R package**