Distribution-free modelling: Quantile GAMs

Matteo Fasiolo (University of Bristol, UK)

matteo.fasiolo@bristol.ac.uk

June 27, 2018

Intro to Generalized Additive Models (GAMs)

Structure:

1. What is quantile regression
2. When is it useful
3. Quantile regression using `qgam`

Structure of the talk

Structure:

1. **What is quantile regression?**
2. When is it useful
3. Quantile regression using `qgam`

What is quantile regression

Regression setting:

- y is our response or dependent variable
- $\mathbf{x}$ is a vector of covariates or independent variables

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

Model is $p_m\{y|\theta_1(\mathbf{x}), \dots, \theta_q(\mathbf{x})\}$, where $\theta_1(\mathbf{x}), \dots, \theta_q(\mathbf{x})$ are parameters.

In a Gaussian model, the mean and/or variance depend on the covariates

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

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

What is quantile regression

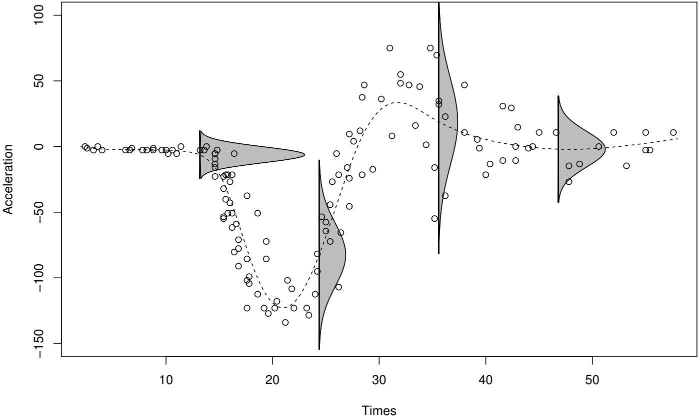


Figure: Gaussian model with variable mean and variance.
In mgcv: `gam(list(y~s(x), ~s(x)), family=gaulss)`.

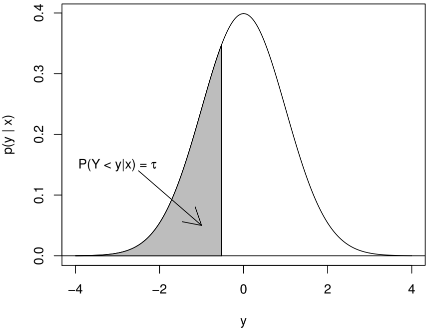
What is quantile regression

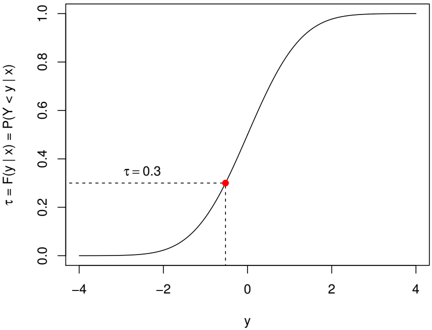
Lots of options for $p_m(y|\mathbf{x})$: binomial, gamma, Poisson, Tweedie...

We consider continuous (not discrete) y.

Let $F(y|\mathbf{x})$ be Prob$(Y \leq y|\mathbf{x})$.

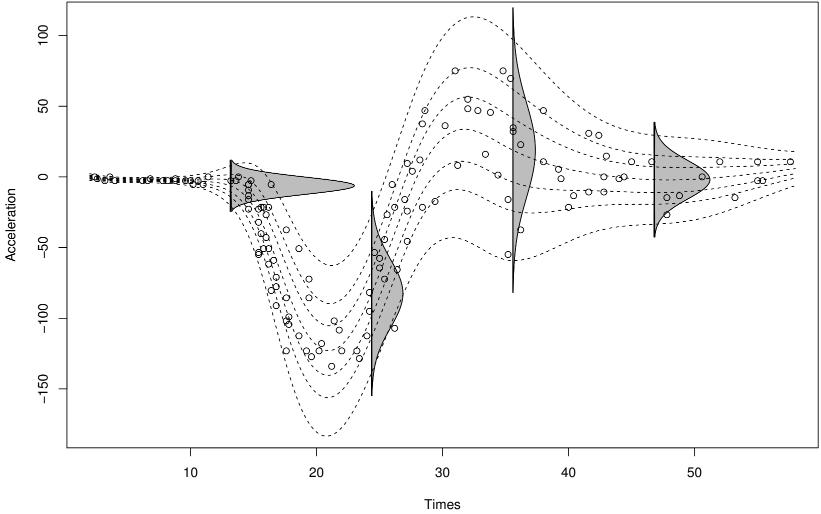
The τ-th ($\tau \in (0,1)$) quantile is $\mu_\tau(\mathbf{x}) = F^{-1}(\tau|\mathbf{x})$.





What is quantile regression

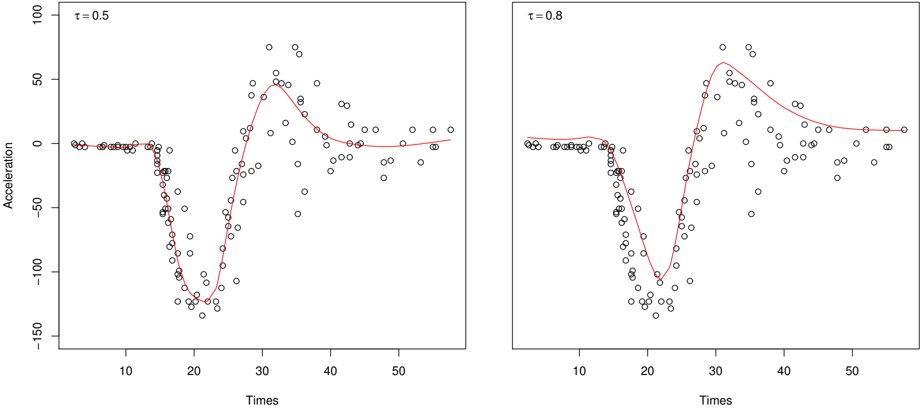
Given $p_m(y|\mathbf{x})$ we can get the conditional quantiles $\mu_\tau(\mathbf{x})$.



What is quantile regression

Quantile regression estimates conditional quantiles $\mu_\tau(\mathbf{x})$ directly.

No model for $p(y|\mathbf{x})$.



What is quantile regression

The τ-th quantile is

$$\mu = F^{-1}(\tau|\mathbf{x}),$$

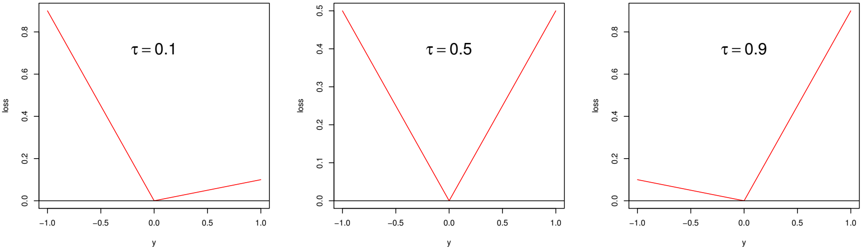
but also the minimizer of

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

where

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

is the "pinball" loss.



What is quantile regression

In **linear quantile regression** $\mu_\tau(\mathbf{x}) = \boldsymbol{\beta}^\mathsf{T}\mathbf{x} = \beta_1 x_1 + \dots \beta_p x_p$.

$\hat{\boldsymbol{\beta}}$ is the minimizer of total pinball loss

$$\hat{\boldsymbol{\beta}} = \underset{\beta}{\text{argmin}}\, L_y(\boldsymbol{\beta}) = \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_i).$$

In **additive quantile regression** $\mu_\tau(\mathbf{x}) = \sum_{j=1}^{m} f_j(\mathbf{x})$.

f_j's can be fixed, random or smooth effects.

$\hat{\boldsymbol{\beta}}$ is the minimizer of total **penalized** pinball loss

$$\hat{\boldsymbol{\beta}} = \underset{\beta}{\text{argmin}} \left\{ L_y(\boldsymbol{\beta}) + \text{Pen}(\boldsymbol{\beta}) \right\}.$$

where $\text{Pen}(\boldsymbol{\beta})$ penalizes the complexity of the f_j's.

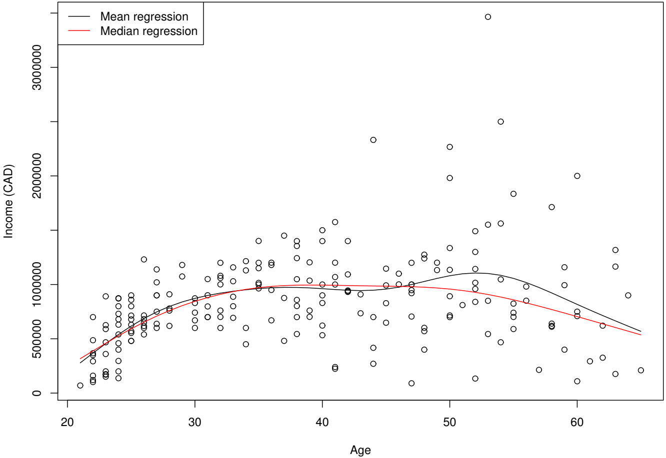
Structure of the talk

Structure:

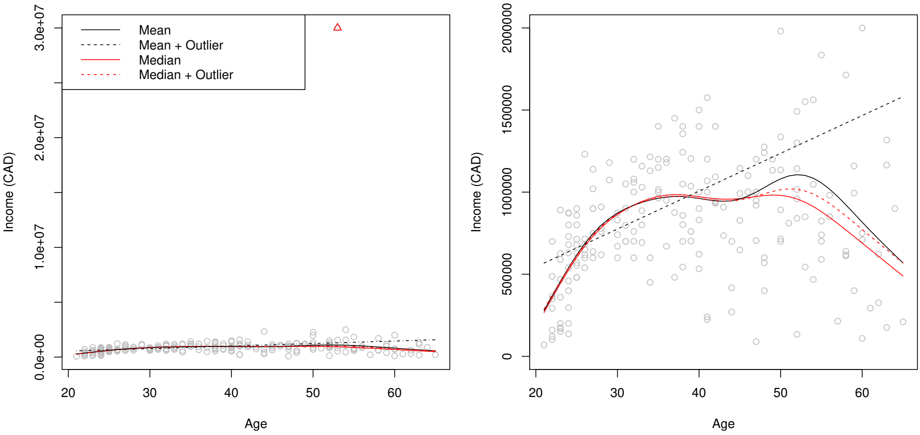
1. What is quantile regression
2. **When is it useful**
3. How to do quantile regression using `qgam`

When is quantile regression useful

Median income is a better indicator of how the "average" person is doing, relative to mean income.



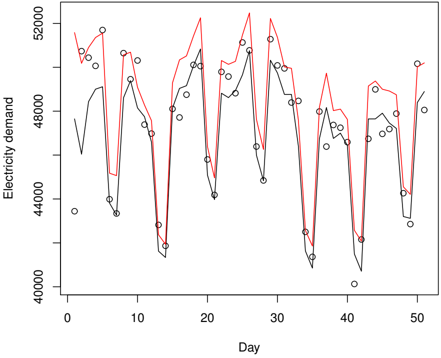
The median is also more **resistant to outliers**.

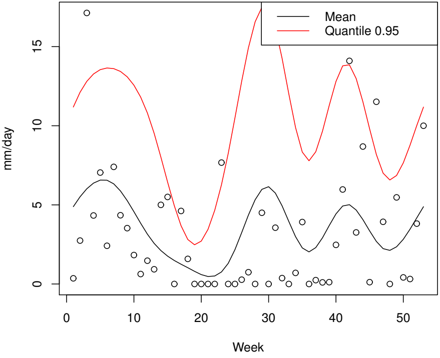


When is quantile regression useful

Some quantiles are more important than others:

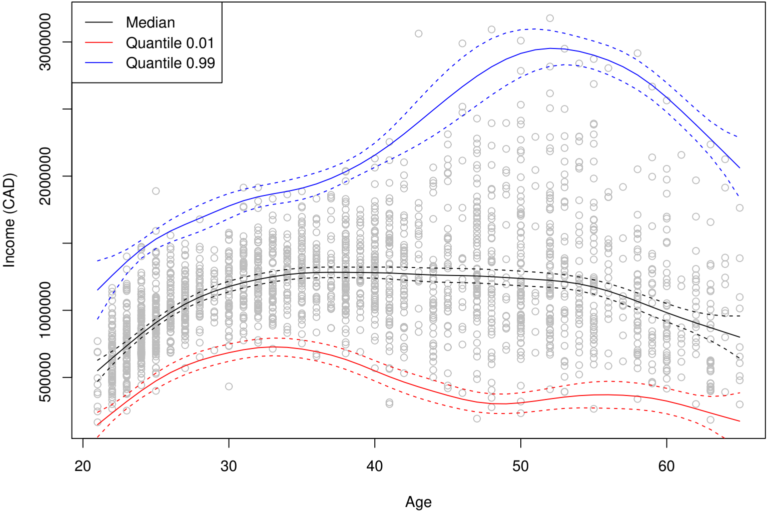
- electricity producers need to satisfy top electricity demand
- urban planners need estimates of extreme rainfall





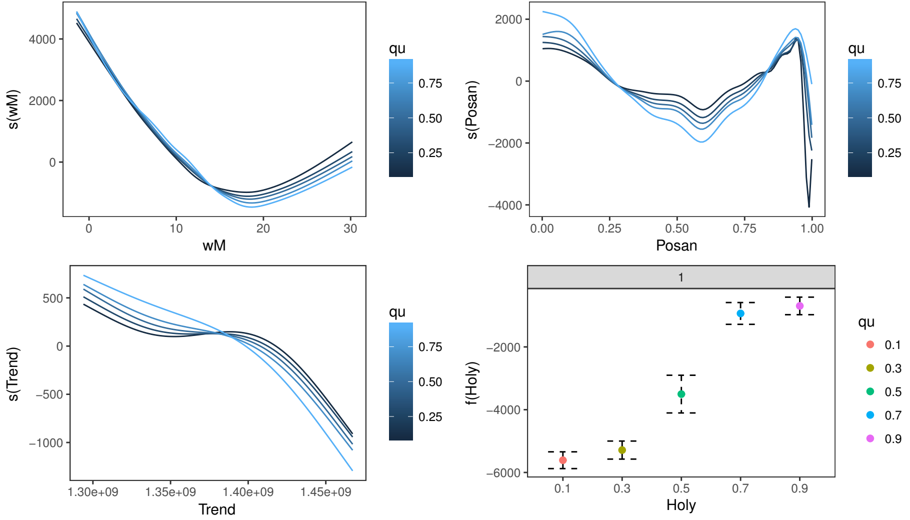
When is quantile regression useful

Effect of explanatory variables may depend on quantile



When is quantile regression useful

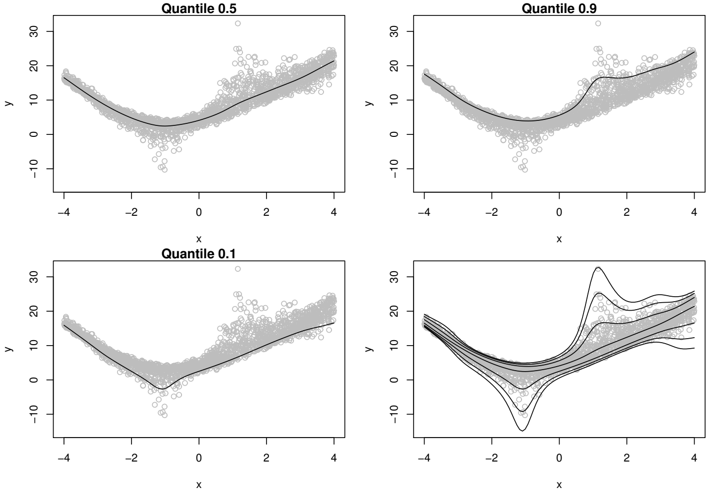
$$q_\tau(\text{Demand}) = f_1(\text{Temp}) + f_2(\text{TimeOfYear}) + f_3(\text{Trend}) + f_4(\text{Holiday}) + \cdots$$



When is quantile regression useful

No assumptions on $p(y|\mathbf{x})$:

- no need to find good model for $p(y|\mathbf{x})$;
- no need to find normalizing transformations (e.g. Box-Cox);



Structure of the talk

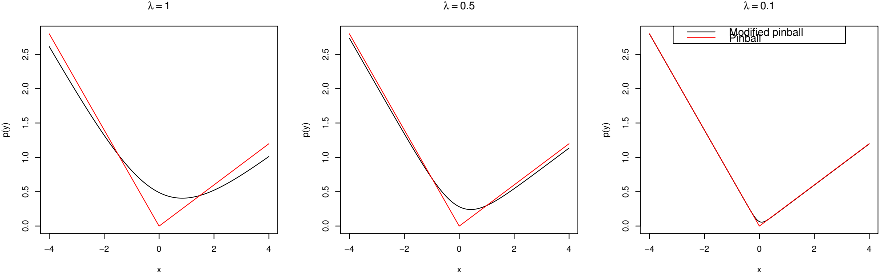
Structure:

1. What is quantile regression?
2. When is it useful
3. **Quantile regression using qgam**

Smoothing the pinball loss

qgam uses a modified loss which we call Extended log-F (ELF) loss.

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



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

Smoothing the pinball loss

Increasing `err` leads to:

- faster and more stable computation
- more bias

Interpretation, if μ^* is minimizer of ELF loss:

$$|F(\mu^*) - \tau| \leq \text{err}$$

`err` is an upper bound on the bias.

By default:

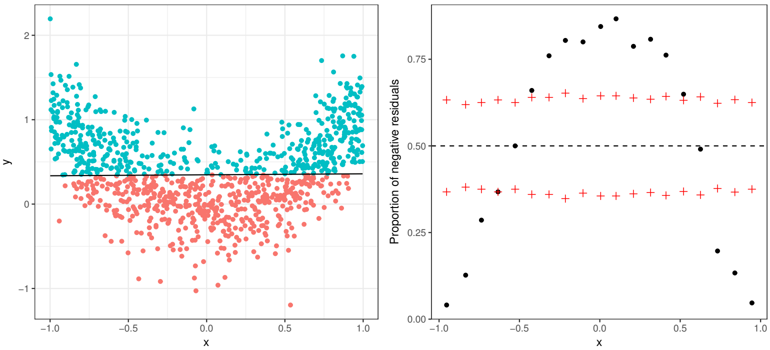
```
qgam(..., err = 0.05, ...)
```

Residual checking

We have no model for $p(y|\mathbf{x})$ → QQ-plots are useless.

We can check the proportion of residuals < 0, which should be $\approx \tau$.

```
check1D(b, "x") + l_gridQCheck1D(qu = 0.5)
```



Model selection

In probabilistic regression we can use Akaike Information Criterion (AIC):

$$\text{AIC} = \underbrace{-2 \log p(\mathbf{y}|\boldsymbol{\beta})}_{\text{goodness of fit}} \quad + \quad \underbrace{2p}_{\text{model complexity}}$$

If $\text{AIC}_{m1} < \text{AIC}_{m2}$ choose model 1.

In quantile regression pinball loss substitutes likelihood $\log p(\mathbf{y}|\boldsymbol{\beta})$.

Maybe justifiable for median regression ($\tau = 0.5$).

Practical approach: choose model with lowest AIC at median and use it for other quantiles.

Probably better: choose model on mean model and use it for quant reg.

THANK YOU!

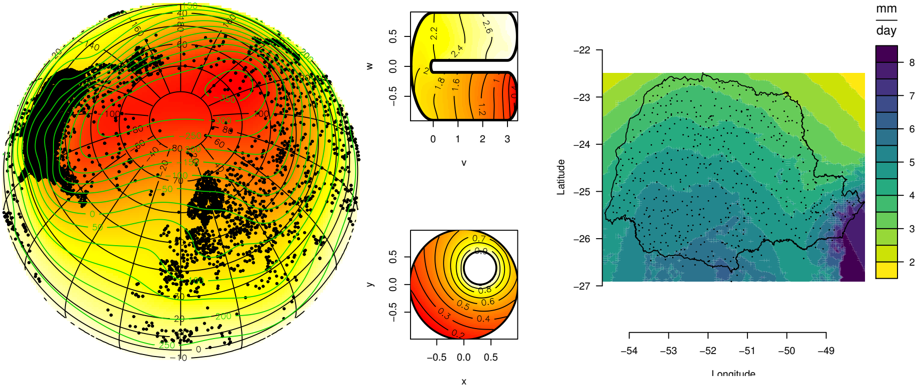


Figure: Examples of quantile GAMs from Fasiolo et al. (2017).

References I

Fasiolo, M., Y. Goude, R. Nedellec, and S. N. Wood (2017). Fast calibrated additive quantile regression. *arXiv preprint arXiv:1707.03307*.