

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.

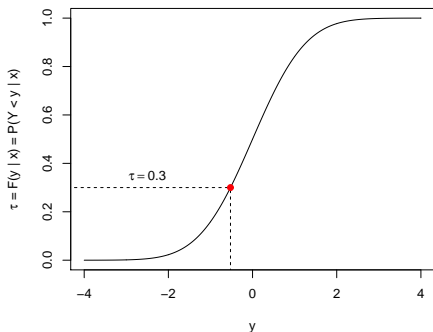
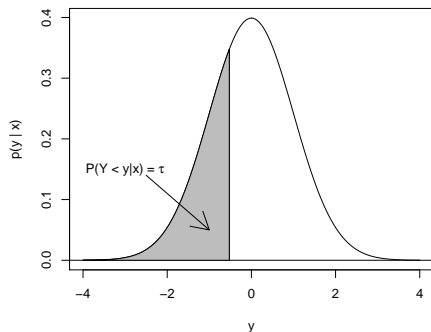
What is quantile regression

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

We consider continuous (not discrete) y .

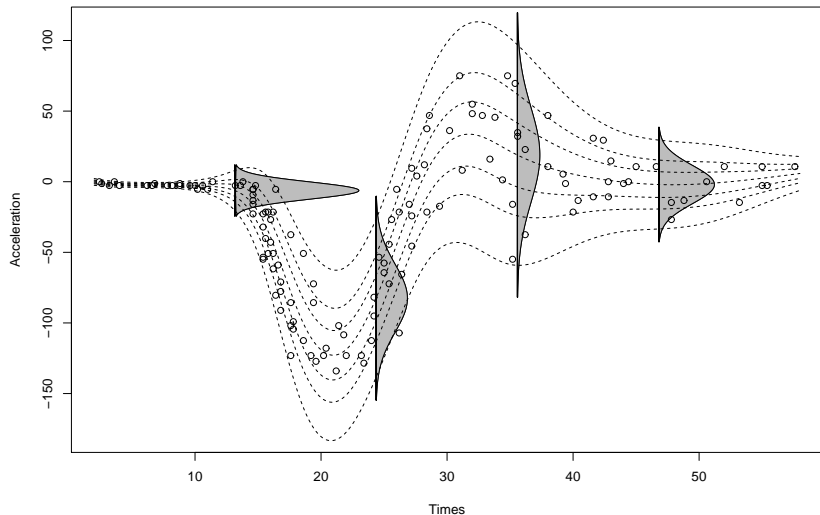
Define $F(y|\mathbf{x}) = \text{Prob}(Y \leq y|\mathbf{x})$.

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



What is quantile regression

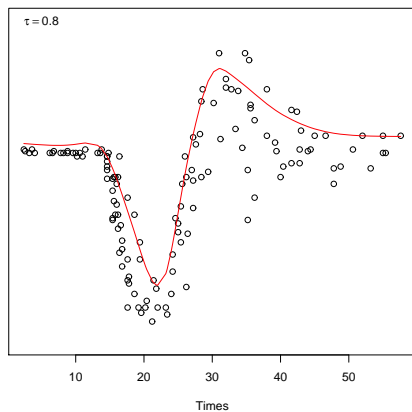
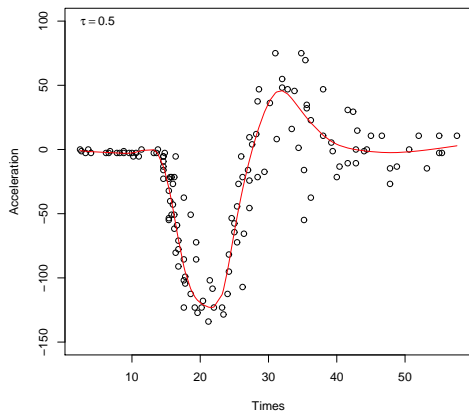
Given $p_m(y|\mathbf{x})$ we can get the conditional quantiles $\mu_\tau(\mathbf{x})$.



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

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

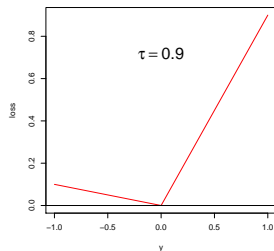
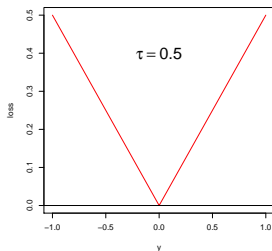
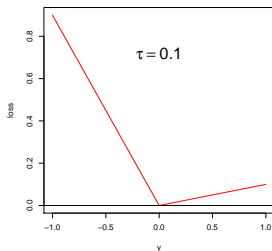
but also the minimizer of

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

where

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

is the “pinball” loss (Koenker, 2005).



What is quantile regression

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

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

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

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

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

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

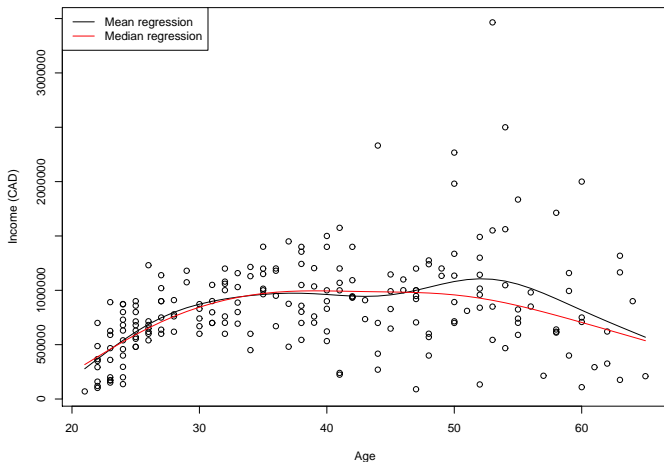
Structure of the talk

Structure:

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

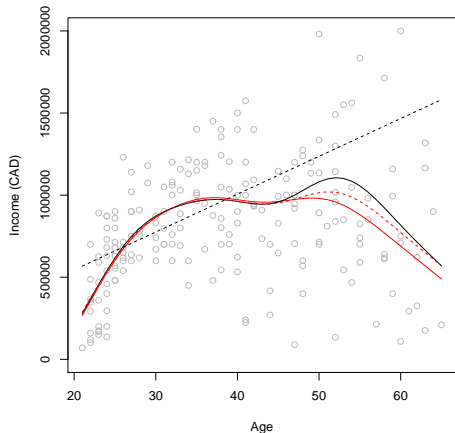
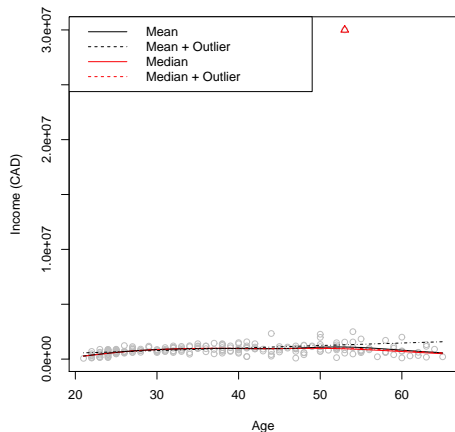
When is quantile regression useful

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



When is quantile regression useful

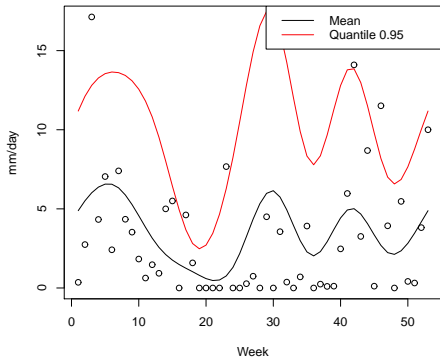
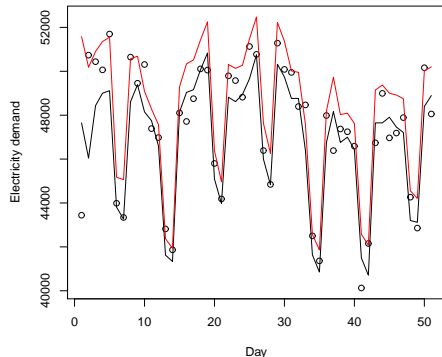
The median is also more **resistant to outliers**.



When is quantile regression useful

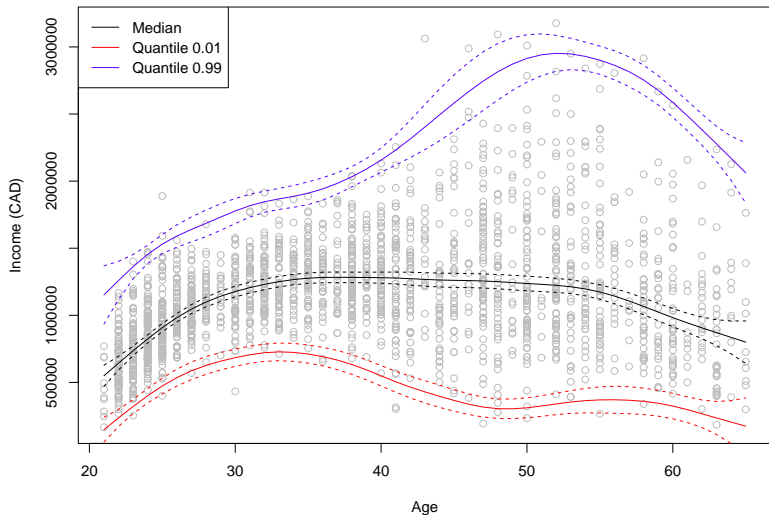
Some quantiles are more important than others:

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



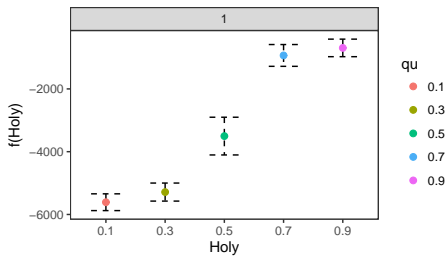
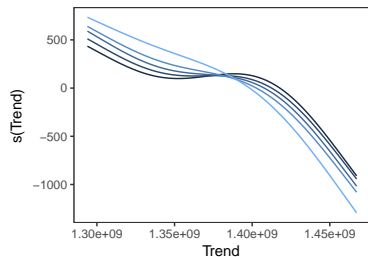
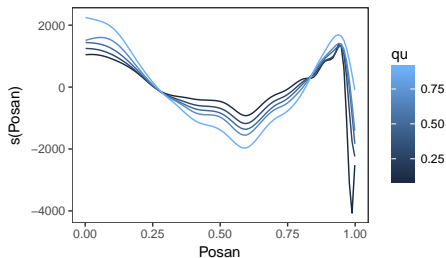
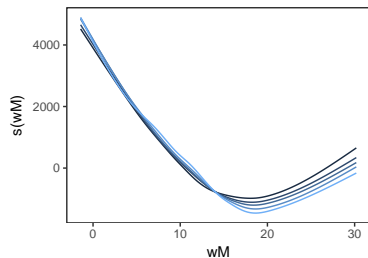
When is quantile regression useful

Effect of explanatory variables may depend on quantile



When is quantile regression useful

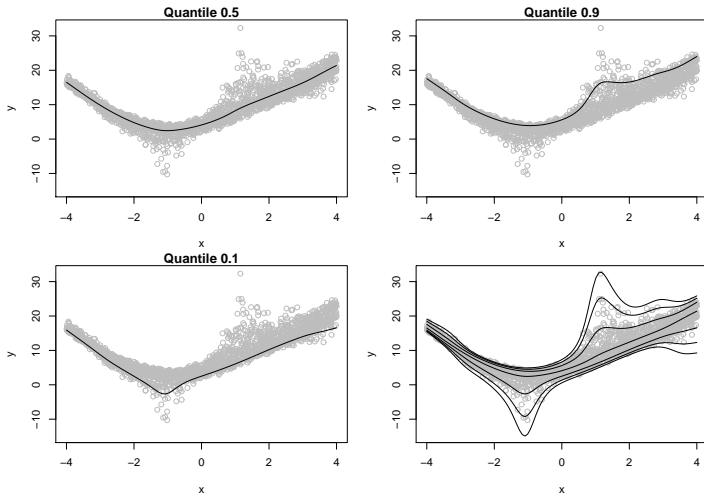
$$q_{\tau}(\text{Demand}) = f_1(\text{Temp}) + f_2(\text{TimeOfYear}) + f_3(\text{Trend}) + f_4(\text{Holiday}) + \dots$$



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

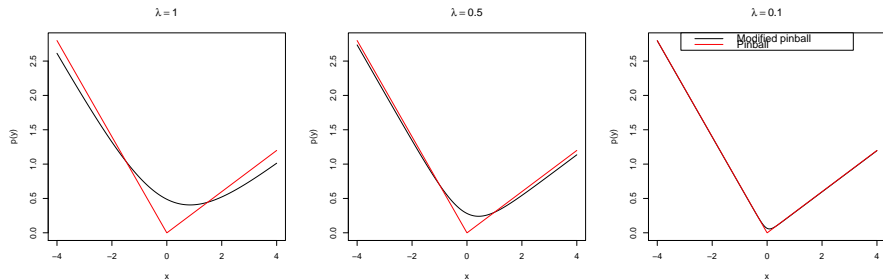
Structure:

- 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 $\lambda \rightarrow 0$, we have recover pinball loss.



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

Smoothing the pinball loss

Increasing `err` leads to:

- faster and more stable computation
- more bias

By default:

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

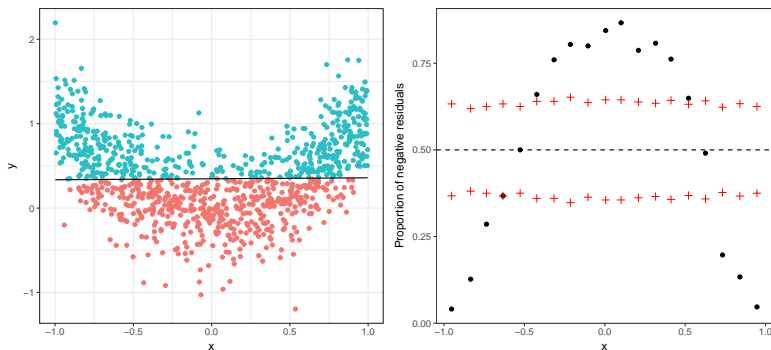
which is a compromise between bias and speed.

Residual checking

We have no model for $p(y|\mathbf{x}) \rightarrow$ 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)
```



THANK YOU!

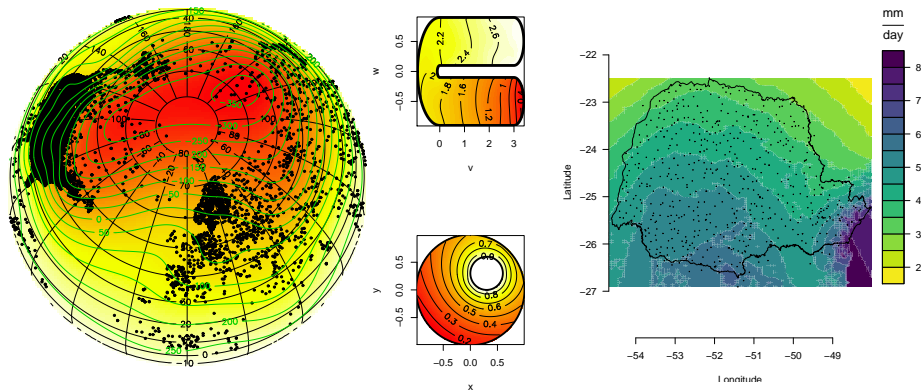


Figure : Examples of quantile GAMs from Fasiolo et al. (2017).

References I

- Fasiolo, M., Y. Goude, R. Nedellec, and S. N. Wood (2017). Fast calibrated additive quantile regression. *arXiv preprint arXiv:1707.03307*.
- Koenker, R. (2005). *Quantile regression*. Number 38. Cambridge university press.