Quantile GAM modelling with `qgam`

Matteo Fasiolo

matteo.fasiolo@bristol.ac.uk

Material available at:

https://github.com/mfasiolo/workshop_BOZEN19

These slides cover:

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.

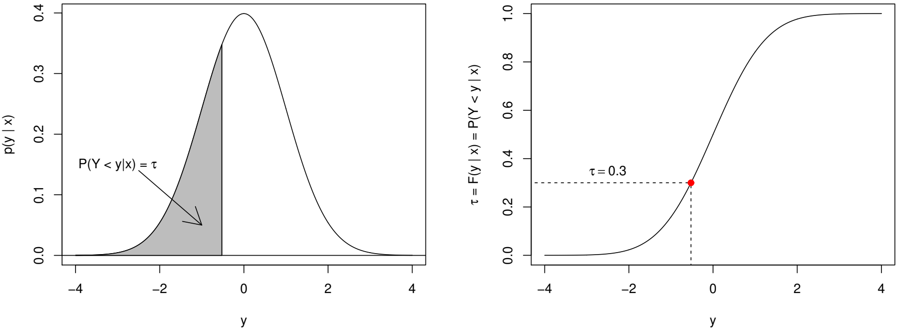
What is quantile regression

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

We consider continuous (not discrete) y.

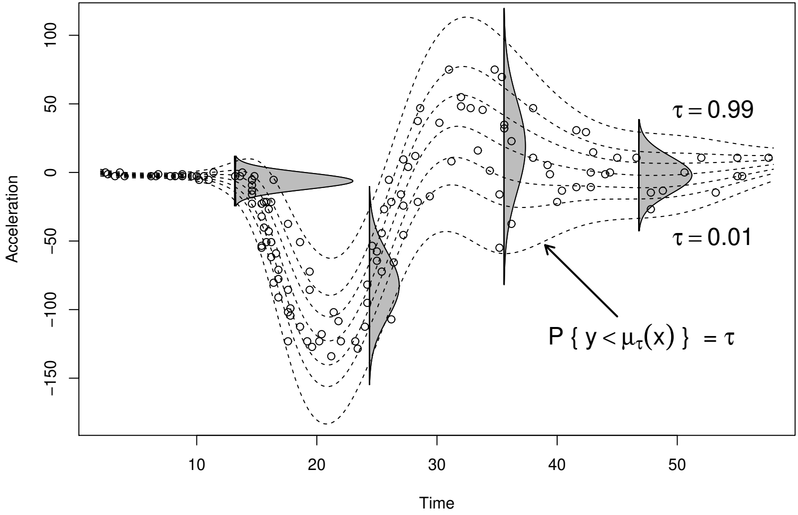
Define $F(y|\mathbf{x}) = \text{Prob}(Y \leq y|\mathbf{x})$.

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



What is quantile regression

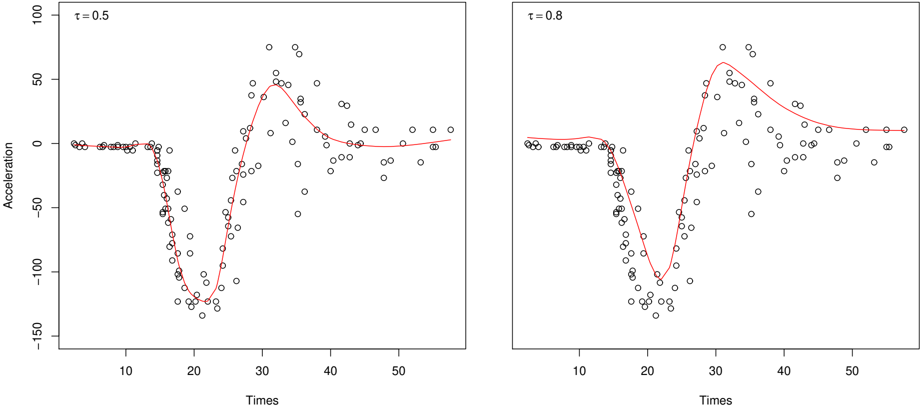
Given $p_m(y|\mathbf{x})$ we can get the conditional quantiles $\mu_\tau(\mathbf{x})$.



What is quantile regression

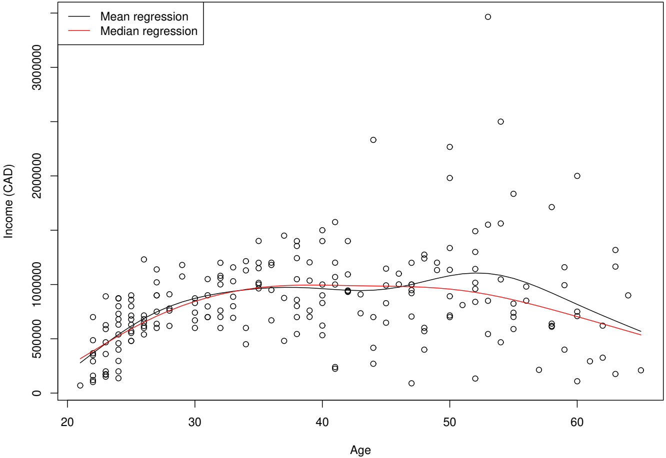
Quantile regression estimates conditional quantiles $\mu_\tau(\mathbf{x})$ directly.

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



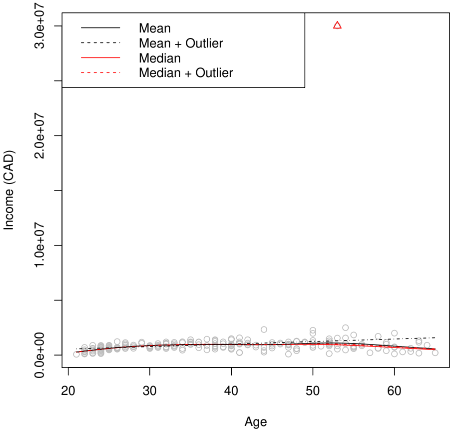
When is quantile regression useful

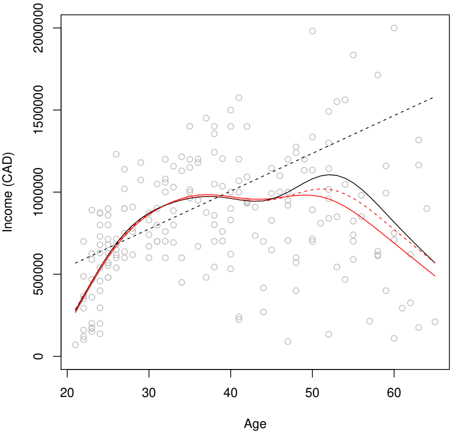
Median income is a better indicator of how the “average” person is doing, relative to mean income.



When is quantile regression useful

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

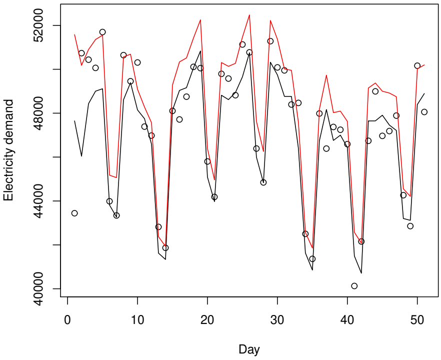


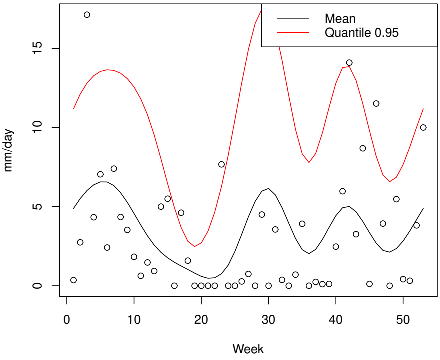


When is quantile regression useful

Some quantiles are more important than others:

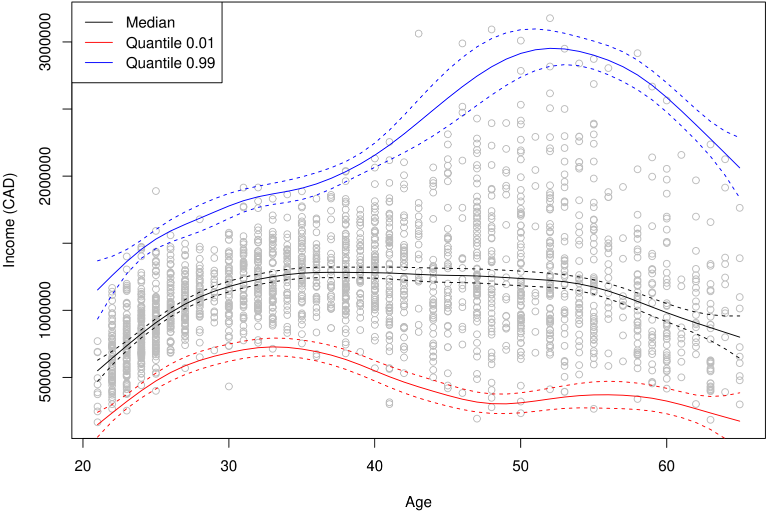
- electricity producers need to satisfy top electricity demand
- urban planners need estimates of extreme rainfall



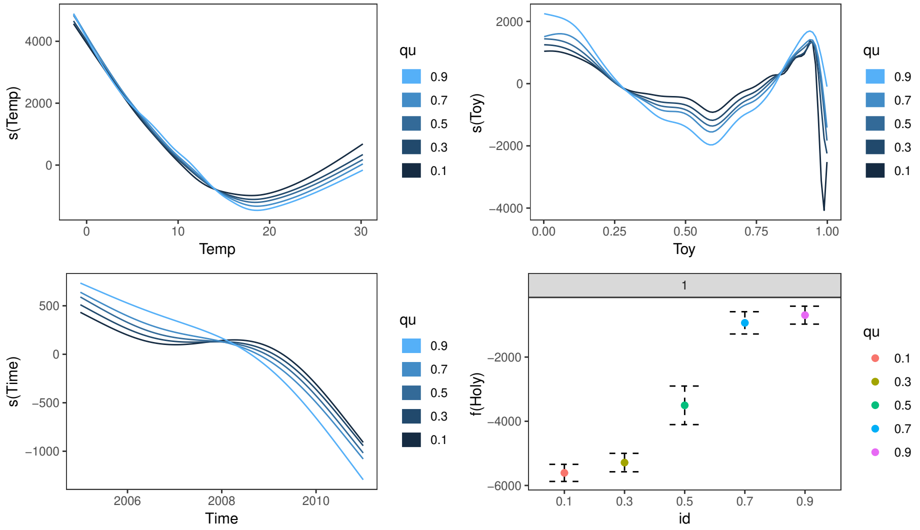


When is quantile regression useful

Effect of explanatory variables may depend on quantile



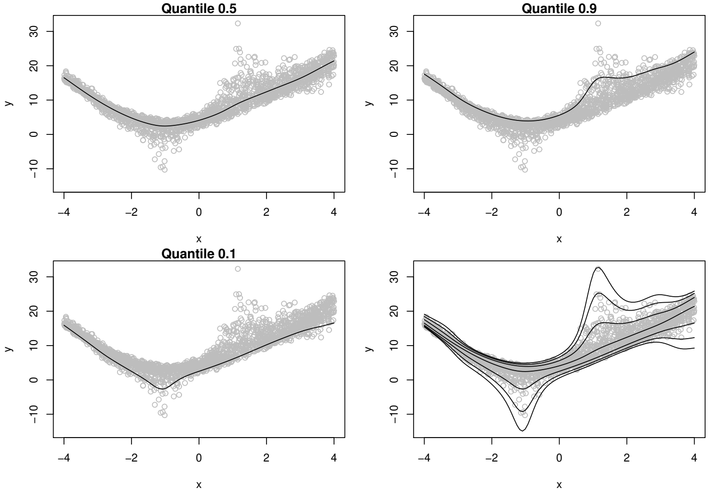
$$q_\tau(\text{Demand}) = f_1(\text{Temp}) + f_2(\text{TimeOfYear}) + f_3(\text{Trend}) + f_4(\text{Holiday}) + \cdots$$



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);



These slides cover:

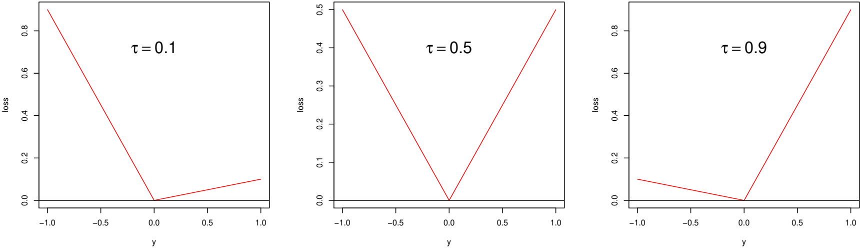
Quantile GAM estimation

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

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

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

where ρ_τ is the "pinball" loss (Koenker, 2005):



In additive modelling context $\mu_\tau(\mathbf{x}) = \sum_{j=1}^m f_j(\mathbf{x}) = \mu_\tau(\boldsymbol{\beta})$.

Quantile GAM estimation

Problem: how to perform Bayesian update $p(\boldsymbol{\beta}|y) \propto p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})$?

A coherent approach is the Gibbs posterior (Bissiri et al., 2016)

$$p(\boldsymbol{\beta}|y) \propto \underbrace{e^{-\frac{1}{\sigma}\rho_\tau\{y-\mu(\boldsymbol{\beta})\}}}_{\text{pseudo } p(y|\boldsymbol{\beta})} p(\boldsymbol{\beta}),$$

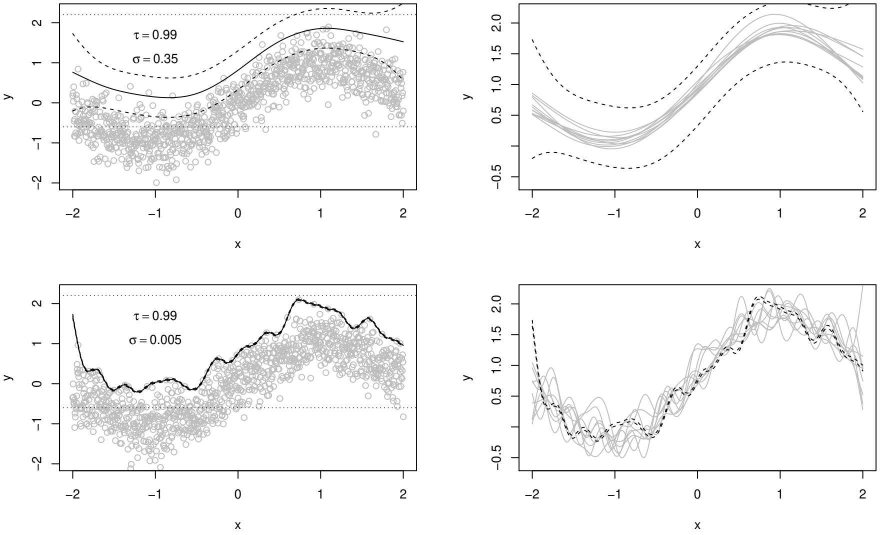
where $1/\sigma > 0$ is the "learning rate".

Recall that $p(\boldsymbol{\beta}) = p(\boldsymbol{\beta}|\boldsymbol{\gamma})$, hence we need to:

- select learning rate $1/\sigma$
- select smoothing parameters $\boldsymbol{\gamma}$
- estimate regression coefficients $\boldsymbol{\beta}$

Technical challenges

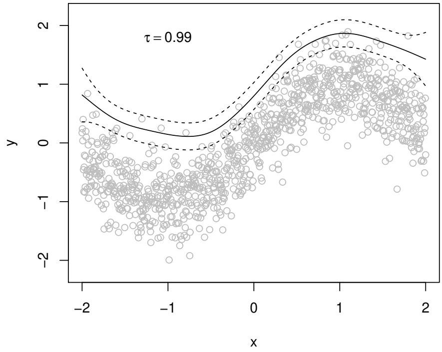
$\boldsymbol{\sigma}$ controls width of credible intervals:

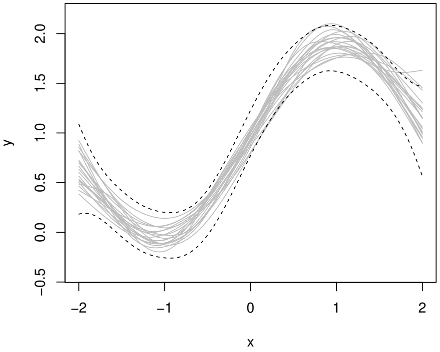


Selecting the learning rate

We select σ so that the model-based uncertainty estimates match the sampling uncertainty, that is σ minimizes

$$\text{CalibrLoss}(\sigma) = \int \text{Dist}\{\text{var}_m(\mathbf{x}), \text{var}_{\mathbb{P}}(\mathbf{x})\} p(\mathbf{x}) d\mathbf{x},$$





Quantile GAM estimation

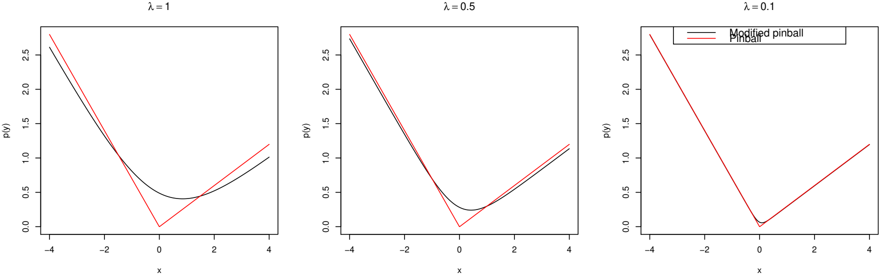
We use a hierarchical fitting framework:

1. Select σ to optimise coverage
$$\hat{\sigma} = \underset{\sigma}{\text{argmin}}\, \text{CalibrLoss}(\sigma).$$
2. For fixed σ, select γ to determine smoothness
$$\hat{\gamma} = \underset{\gamma}{\text{argmax}}\, \text{LAML}(\gamma)$$
where $\text{LAML}(\gamma) \approx p(y|\gamma) = \int p(y, \boldsymbol{\beta}|\gamma) = \int p(y|\boldsymbol{\beta})p(\boldsymbol{\beta}|\gamma)d\boldsymbol{\beta}$.
3. For fixed γ and σ, estimate $\boldsymbol{\beta}$
$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmax}}\ \log p(\boldsymbol{\beta}|y).$$

Quantile GAM estimation

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

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

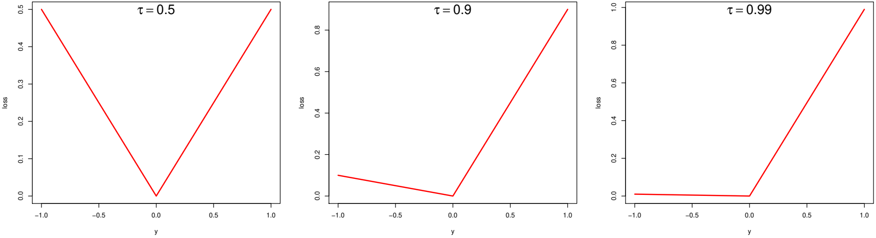


Since `qgam` 1.3.0, λ (`err` parameter) is selected automatically.

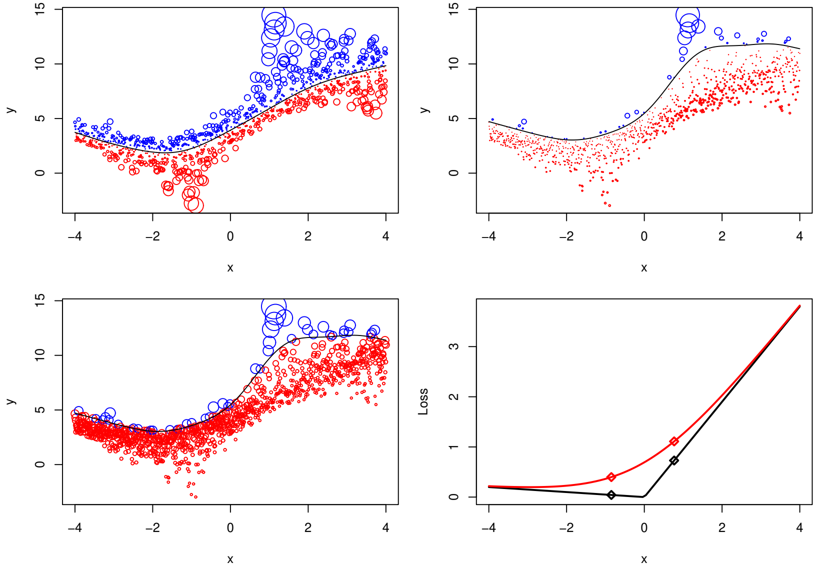
Selecting the learning rate

Motivation for using ELF:

pinball loss becomes very asymmetric on extreme quantiles.



Smoothing the pinball loss



λ (called `err` in `qgam`) selected to **balance variance and bias**.

These slides cover:

Demonstration in R

For more details on methodology, see:

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

and the file "intro_to_qgam.pdf".

For more software training material see

`http://mfasiolo.github.io/qgam/articles/qgam.html`

`https://mfasiolo.github.io/mgcViz/articles/qgam_mgcViz.html`

THANK YOU!

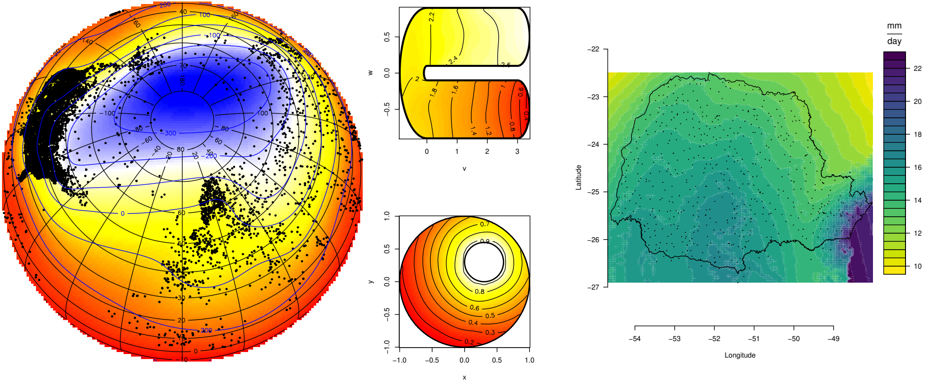


Figure: Examples of quantile GAMs from Fasiolo et al. (2017).

References I

Bissiri, P. G., C. Holmes, and S. G. Walker (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

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.