Quantile GAM modelling with qgam

Matteo Fasiolo

matteo.fasiolo@bristol.ac.uk

Material available at:

https://github.com/mfasiolo/GAM_Workshop_Enbis_EDF_23

These slides cover:

1. Intro to quantile GAM models
2. Fitting GAMs with mgcv
3. Fitting GAMs with qgam
4. Quantile GAM modelling with qgam

What is quantile regression

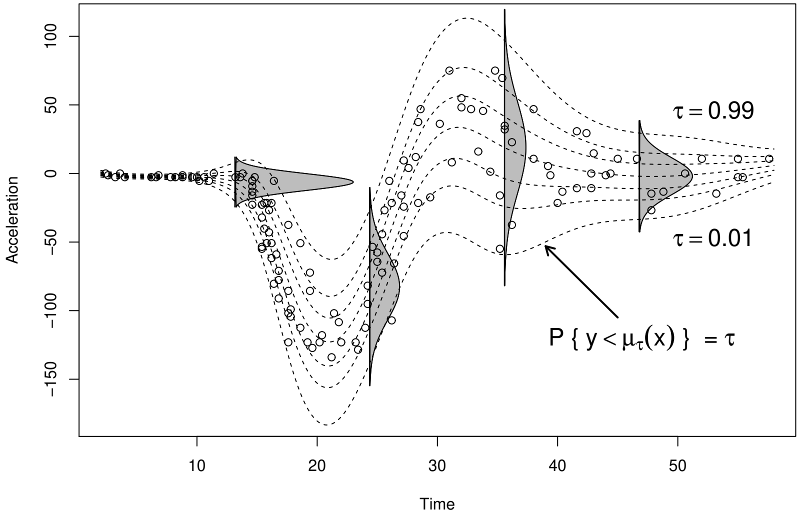
Regression setting:

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

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

Model is $\text{Distr}_m\{y|\theta_1(\boldsymbol{x}), \dots, \theta_q(\boldsymbol{x})\}$, where $\theta_1(\boldsymbol{x}), \dots, \theta_q(\boldsymbol{x})$ are parameters.

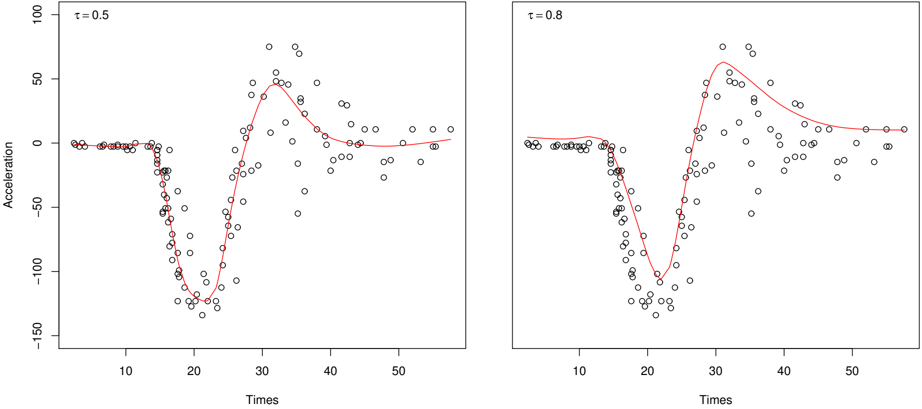
Given $\text{Distr}_m(y|\boldsymbol{x})$ we can get the conditional quantiles $\mu_\tau(\boldsymbol{x})$.



What is quantile regression

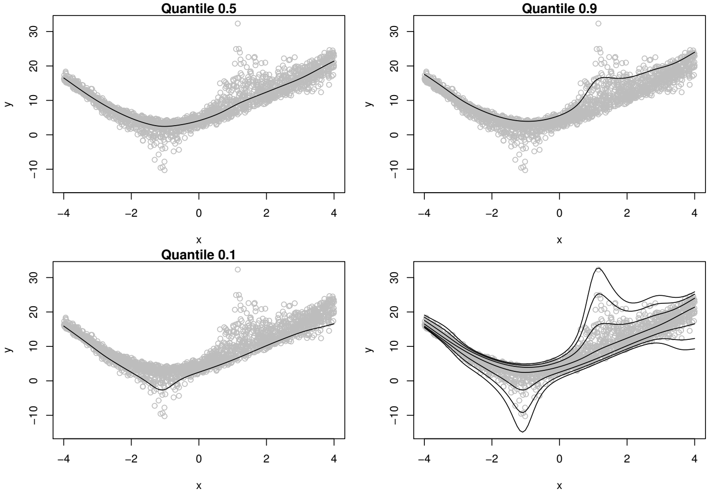
Quantile regression estimates conditional quantiles $\mu_\tau(\boldsymbol{x})$ directly.

No model for $\text{Distr}(y|\boldsymbol{x})$.

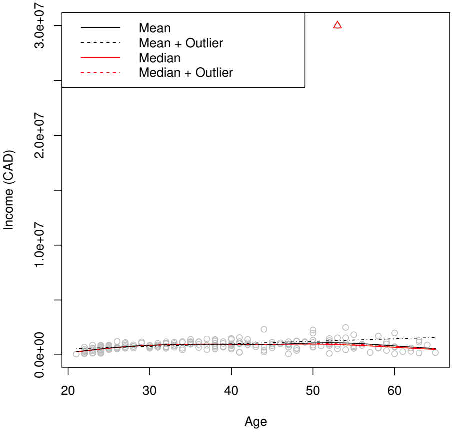


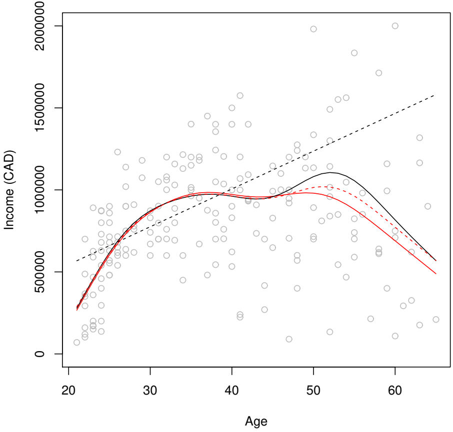
No assumptions on Distr$(y|\boldsymbol{x})$:

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



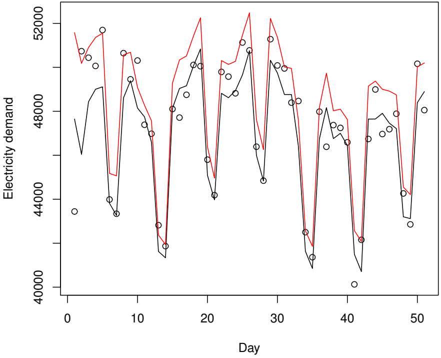
The median is also more **resistant to outliers**.

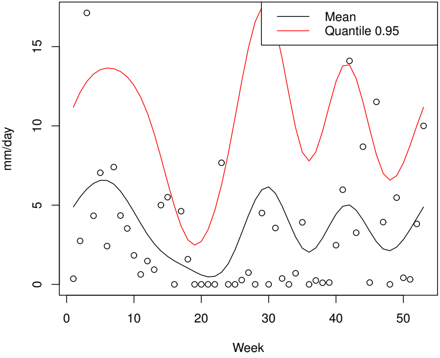




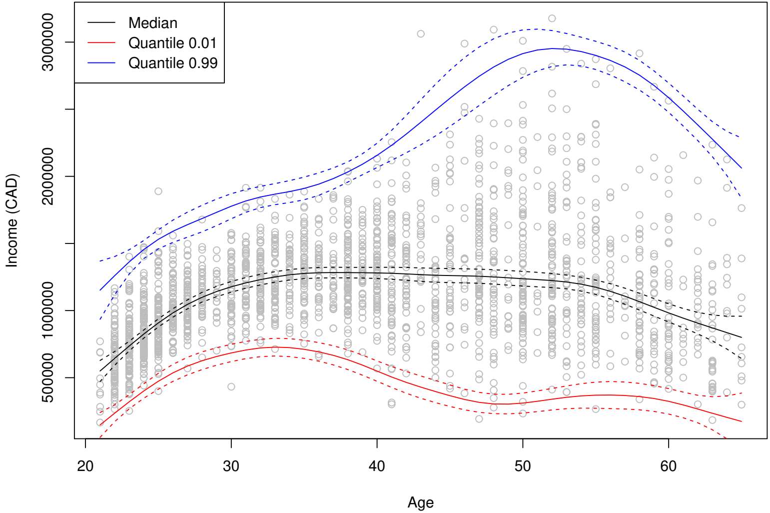
Some quantiles are more important than others:

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





Effect of explanatory variables may depend on quantile



These slides cover:

1. Intro to quantile GAM models
2. Fitting GAMs with mgcv
3. Fitting GAMs with qgam
4. Quantile GAM modelling with qgam

Model fitting

Recall the GAM model structure:

$$y|\boldsymbol{x} \sim \text{Distr}\{y|\mu(\boldsymbol{x}), \boldsymbol{\theta}\} \quad \text{where} \quad g(\mu(\boldsymbol{x})) = \sum_{j=1}^{m} f_j(\boldsymbol{x}).$$

In `mgcv` $\boldsymbol{\beta}$ estimated by maximising **penalized** log-likelihood

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmax}}\ \text{PenLogLik}(\boldsymbol{\beta}|\gamma) = \underset{\boldsymbol{\beta}}{\text{argmax}} \Big\{ \overbrace{\log p(\boldsymbol{y}|\boldsymbol{\beta})}^{\text{goodness of fit}} - \underbrace{\text{Pen}(\boldsymbol{\beta}|\gamma)}_{\text{penalize complexity}} \Big\}$$

where:

- $\log p(\boldsymbol{y}|\boldsymbol{\beta})$ is log-likelihood
- $\text{Pen}(\boldsymbol{\beta}|\gamma)$ penalizes the complexity of the f_j's
- $\gamma > 0$ smoothing parameters ($\uparrow \gamma \uparrow$smoothness)

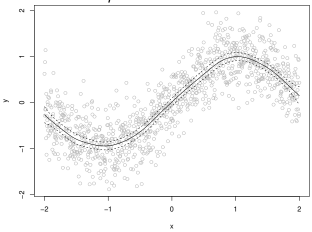
`mgcv` uses a hierarchical fitting framework:

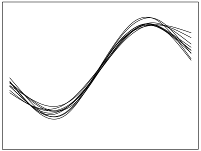
1. Select γ to determine smoothness

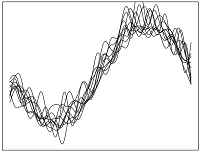
$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmax}}\ \text{LAML}(\gamma).$$

2. For fixed γ, estimate $\boldsymbol{\beta}$ to determine actual fit

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\ \text{PenLogLik}(\boldsymbol{\beta}|\gamma).$$



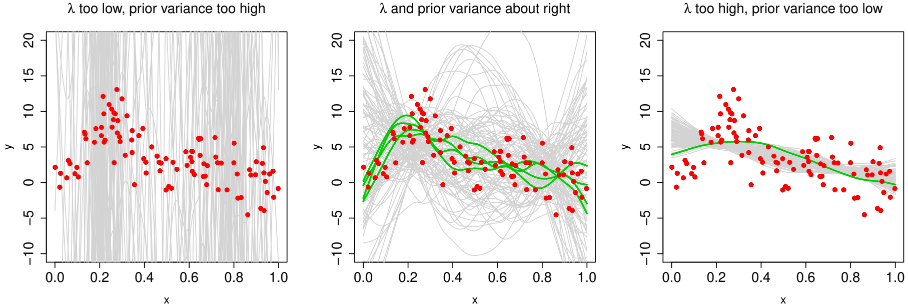




What is the Laplace Approximate Marginal Likelihood?

Under Bayesian perspective, let $p(\boldsymbol{\beta}|\gamma)$ be prior on $\boldsymbol{\beta}$ and

$$\text{LAML}(\gamma) \approx p(\mathbf{y}|\gamma) = \int p(\mathbf{y}|\boldsymbol{\beta})p(\boldsymbol{\beta}|\gamma)d\boldsymbol{\beta}.$$



(In plots above λ should be γ)

Alternatives LAML for γ selection:

- Generalized Cross-Validation (GCV)
- Akaike Infomation Criterion (AIC)

but LAML is most widely applicable in `mgcv`.

To choose γ estimation method in `mgcv`

```
fit <- gam(y ~ ..., method = "REML")
```

see `?gam`.

LAML is the default for multi-parameter GAMs.

These slides cover:

1. Intro to quantile GAM models
2. Fitting GAMs with mgcv
3. Fitting GAMs with qgam
4. Quantile GAM modelling with qgam

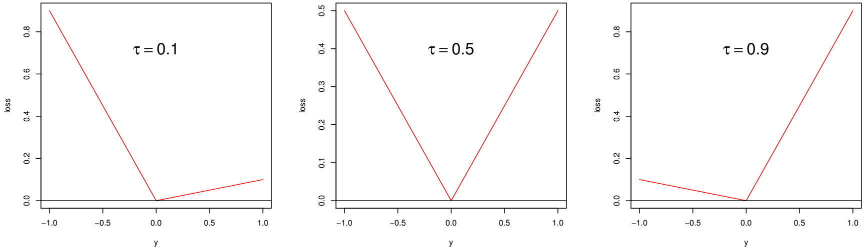
Quantile GAM fitting

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

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

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

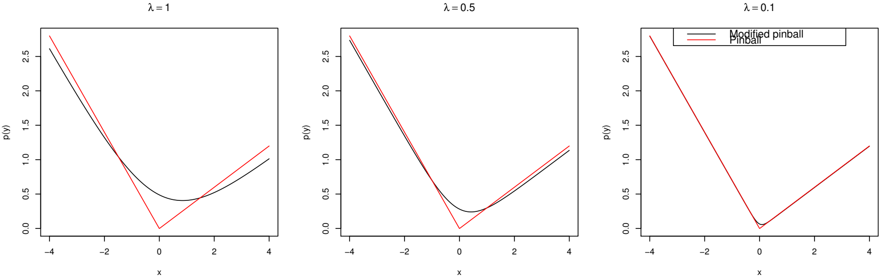
where ρ_τ is the "pinball" loss (Koenker, 2005):



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

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

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



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

Smoothing the loss has statistical advantages, see Fasiolo et al. (2021a).

Recall β estimated by minimising negative **penalized** log-likelihood

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \; -\text{PenLogLik}(\boldsymbol{\beta}|\boldsymbol{\gamma}) = \underset{\boldsymbol{\beta}}{\text{argmin}} \left\{ -\overbrace{\log p(\boldsymbol{y}|\boldsymbol{\beta})}^{\text{goodness of fit}} + \underbrace{\text{Pen}(\boldsymbol{\beta}|\boldsymbol{\gamma})}_{\text{penalize complexity}} \right\}.$$

We plug the ELF loss in place of $-\log p(\boldsymbol{y}|\boldsymbol{\beta})$ so

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \; \text{PenElfLoss}(\boldsymbol{\beta}|\boldsymbol{\gamma}) = \underset{\boldsymbol{\beta}}{\text{argmin}} \left\{ \text{ELFLoss}(\boldsymbol{y}|\boldsymbol{\beta}) + \text{Pen}(\boldsymbol{\beta}|\boldsymbol{\gamma}) \right\}.$$

Getting a good fit requires adding a new parameter, the **learning rate** σ.

We use a hierarchical fitting framework:

1. Select σ to optimise coverage

$$\hat{\sigma} = \underset{\sigma}{\operatorname{argmin}} \, \mathsf{CalibrLoss}(\sigma).$$

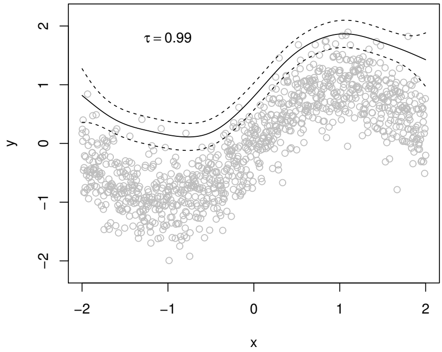
2. For fixed σ, select γ to determine smoothness

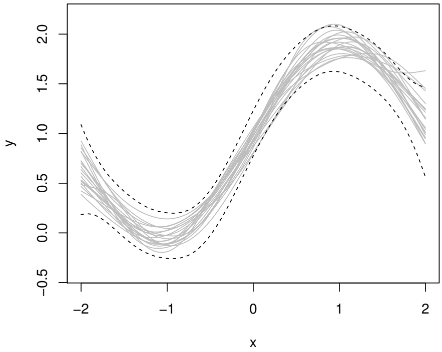
$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmax}} \, \mathsf{LAML}(\gamma).$$

3. For fixed γ and σ, estimate $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \, \mathsf{PenElfLoss}(\boldsymbol{\beta}|\gamma)$$

Minimise CalibrLoss(σ) to match model-based and sampling uncertainty.





NOTE: we can let σ and λ vary with $\boldsymbol{x}$ (see R demo)

These slides cover:

1. Intro to quantile GAM models
2. Fitting GAMs with mgcv
3. Fitting GAMs with qgam
4. Quantile GAM modelling with qgam

Demonstration in R

For more details on methodology, see Fasiolo et al. (2021a) and Fasiolo et al. (2021b).

Ben Griffiths (EDF-sponsored PhD) is working on Big Data (`bam`) method from QGAMs.

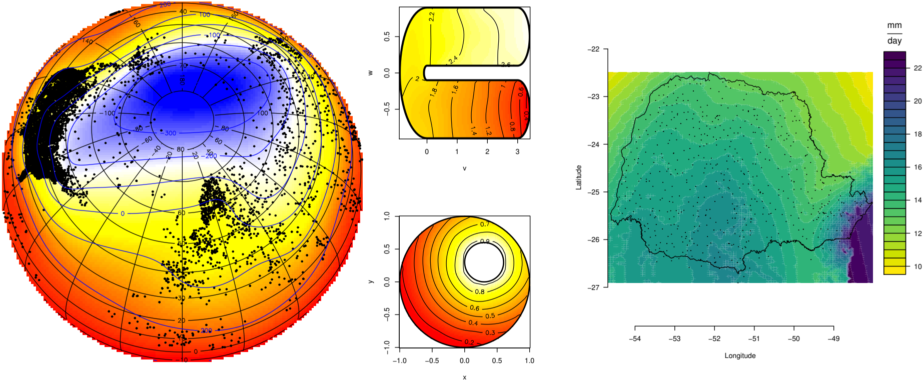
Email him at `ben.griffiths@bristol.ac.uk` to keep updated!

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!



Examples of quantile GAMs from Fasiolo et al. (2021a).

References I

Fasiolo, M., S. N. Wood, M. Zaffran, R. Nedellec, and Y. Goude (2021a). Fast calibrated additive quantile regression. *Journal of the American Statistical Association 116*(535), 1402–1412.

Fasiolo, M., S. N. Wood, M. Zaffran, R. Nedellec, and Y. Goude (2021b). qgam: Bayesian nonparametric quantile regression modeling in r. *Journal of statistical software 100*(9).

Koenker, R. (2005). *Quantile regression*. Number 38. Cambridge university press.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(1), 3–36.

Wood, S. N., N. Pya, and B. Säfken (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association 111*(516), 1548–1575.