Beyond mean modelling: GAMLSS and quantile GAMs

Matteo Fasiolo (University of Bristol, UK)

matteo.fasiolo@bristol.ac.uk

July 11, 2018

GAMLSS

Structure:

1. Intro to GAMs for Location Scale and Shape
2. Intro to quantile GAMs
3. GAMLSS and QGAM modelling in mgcv and qgam

Beyond mean modelling: GAMLSS models

Structure:

1. **Intro to GAMs for Location Scale and Shape**
2. Intro to quantile GAMs
3. GAMLSS and QGAM modelling in mgcv and qgam

Intro to GAMLSS models

Recall GAM model structure:

$$y|\mathbf{x} \sim \text{Distr}\{y|\theta_1 = \mu(\mathbf{x}), \theta_2, \dots, \theta_p\},$$

where

$$\mathbb{E}(y|\mathbf{x}) = \mu(\mathbf{x}) = g^{-1}\Big\{\sum_{j=1}^{m} f_j(\mathbf{x})\Big\},$$

and g is the link function.

Example, Scaled Student-t distribution:

- location $\mu(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$
- scale $\theta_2 = \sigma$
- shape $\theta_3 = \nu$

Intro to GAMLSS models

In Generalized Additive Models for Location Scale and Shape (GAMLSS) we let scale and shape change with the covariates $\mathbf{x}$.

GAMLSS model structure:

$$y|\mathbf{x} \sim \text{Distr}\{y|\theta_1 = \mu_1(\mathbf{x}), \theta_2 = \mu_2(\mathbf{x}), \dots, \theta_p = \mu_p(\mathbf{x})\},$$

where

$$\mu_1(\mathbf{x}) = g_1^{-1}\Big\{\sum_{j=1}^{m} f_j^1(\mathbf{x})\Big\},$$
$$\dots$$
$$\mu_p(\mathbf{x}) = g_p^{-1}\Big\{\sum_{j=1}^{m} f_j^p(\mathbf{x})\Big\},$$

and $g_1, \dots, g_p$ are link function.

Intro to GAMLSS models

Example: **Gaussian model for location and scale**

Model is

$$y|\mathbf{x} \sim N\{y|\mu(\mathbf{x}), \sigma(\mathbf{x})\}$$

where

$$\mathbb{E}(y|\mathbf{x}) = \mu(\mathbf{x}) = \sum_{j=1}^{m} f_j^1(\mathbf{x})$$

$$\text{var}(y|\mathbf{x})^{1/2} = \sigma(\mathbf{x}) = \exp\Big\{\sum_{j=1}^{m} f_j^2(\mathbf{x})\Big\}$$

that is $g_2 = \log$ to guarantee $\sigma > 0$.

Intro to GAMLSS models

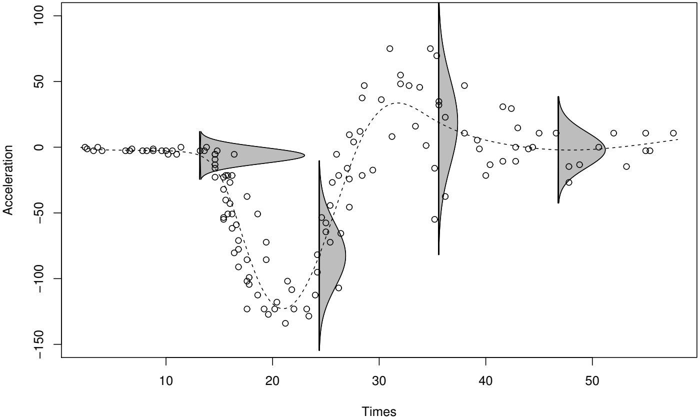
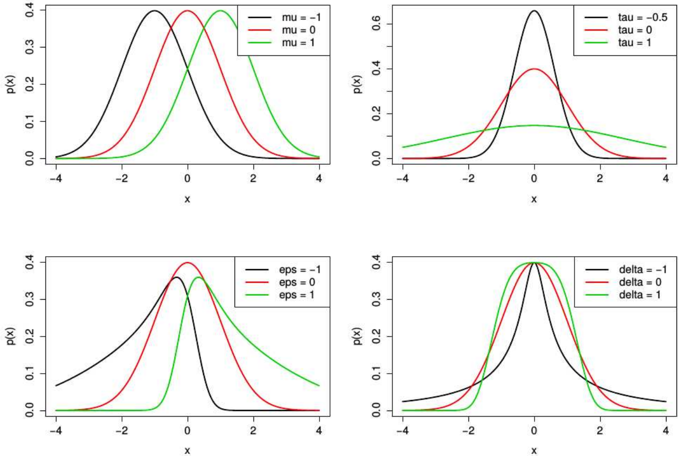


Figure: Gaussian model with variable mean and variance.
In mgcv: `gam(list(y~s(x), ~s(x)), family=gaulss)`.

Example: **Sinh-arcsinh (shash) distribution**

Four parameter distribution where location, scale, skewness (asymmetry) and kurtosis (tail behaviour) can depend on **x**.



Intro to GAMLSS models

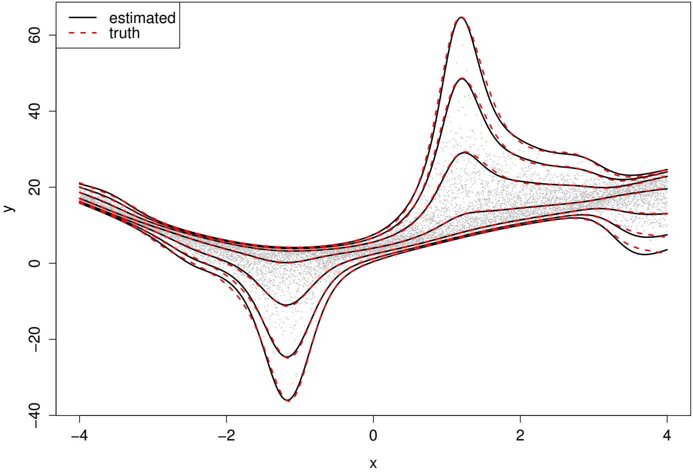
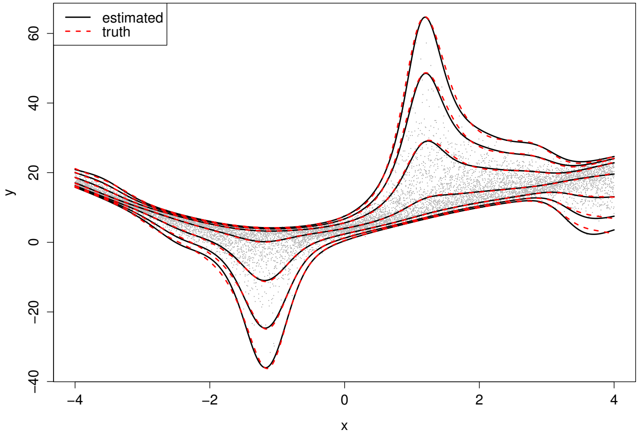


Figure: `gam(list(y~s(x), ~s(x), ~s(x), ~s(x)), family=shash).`

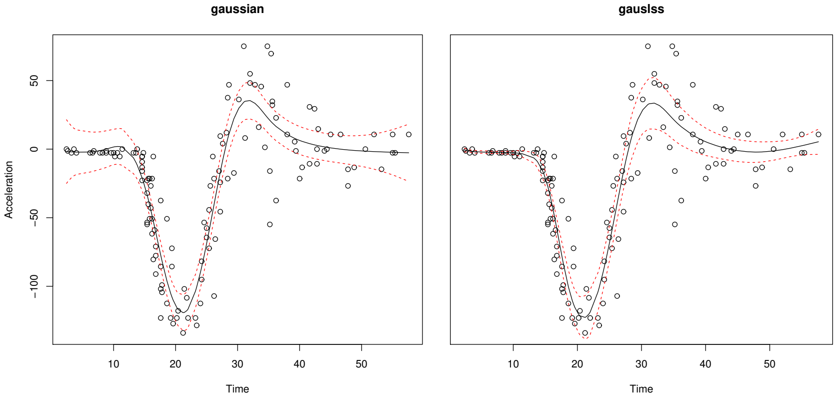
Why is this useful?

R1: you might be interested in whole distribution $y|\mathbf{x}$ not just $\mathbb{E}(y|\mathbf{x})$.



Why is this useful?

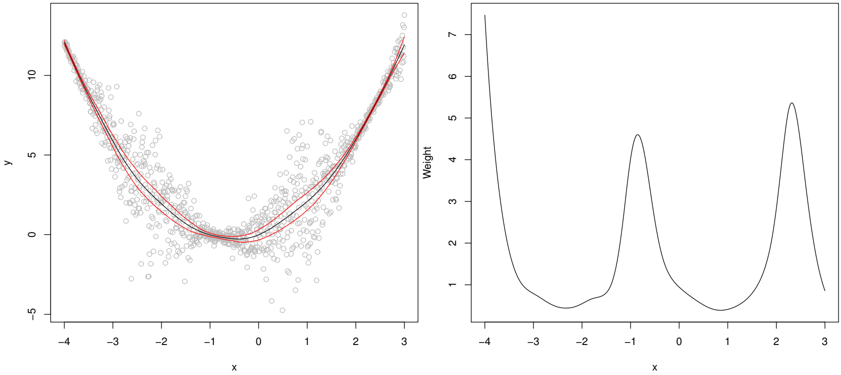
R2: standard GAM inference (e.g. p-value & confidence interval) is valid if the model for $y|\mathbf{x}$ is correct



Intro to GAMLSS models

Why is this useful?

R3: the accuracy of the fit is improved if the weight of each observation is inversely proportional to $\mathrm{Var}(y|\mathbf{x})$.



Beyond mean modelling: GAMLSS models

Structure:

1. Intro to GAMs for Location Scale and Shape
2. **Intro to quantile GAMs**
3. GAMLSS and QGAM modelling in mgcv and 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.

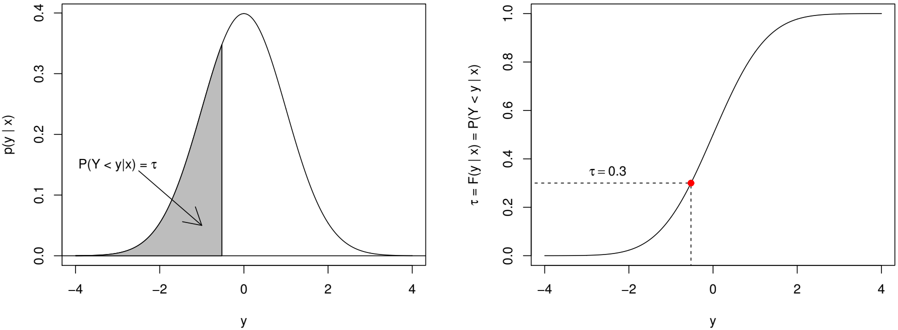
What is quantile regression

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

We consider continuous (not discrete) y.

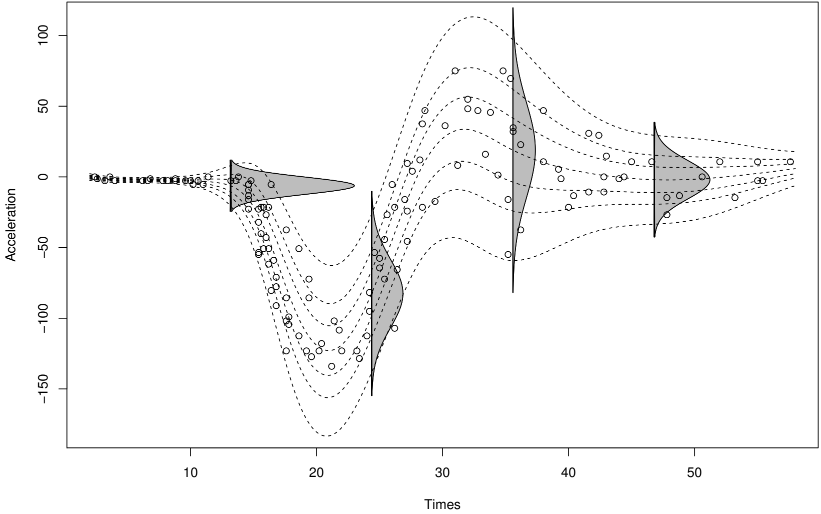
Define $F(y|\mathbf{x}) = \text{Prob}(Y \leq y|\mathbf{x})$.

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



What is quantile regression

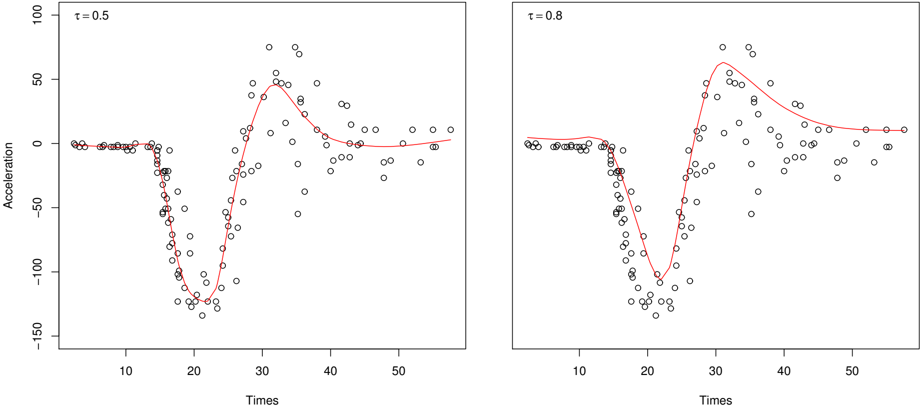
Given $p_m(y|\mathbf{x})$ we can get the conditional quantiles $\mu_\tau(\mathbf{x})$.



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

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

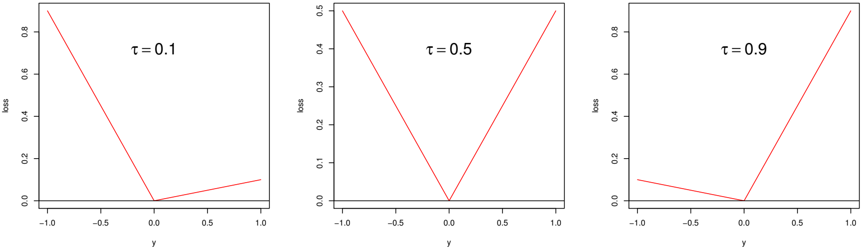
but also the minimizer of

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

where

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

is the "pinball" loss.



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

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

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.

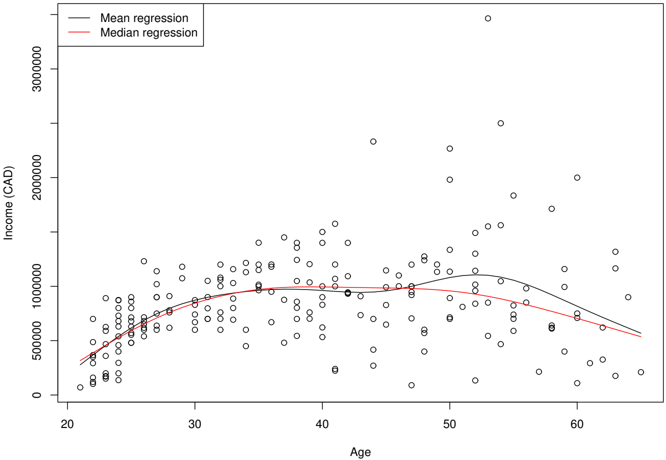
$\hat{\boldsymbol{\beta}}$ is the minimizer of total **penalized** pinball loss

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

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

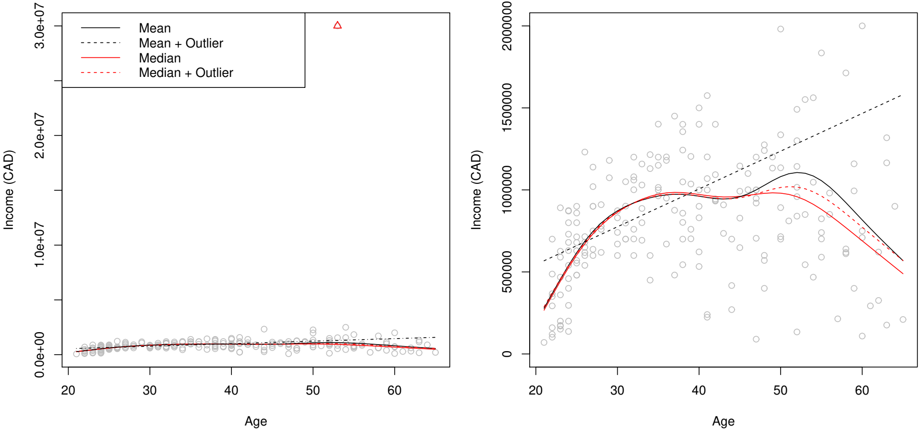
When is quantile regression useful

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



When is quantile regression useful

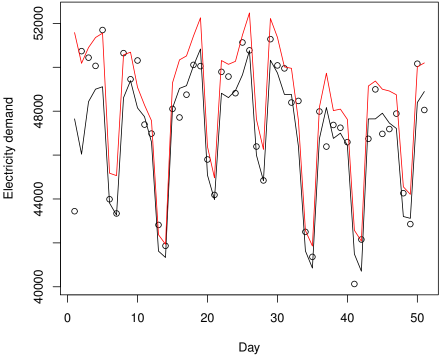
The median is also more **resistant to outliers**.

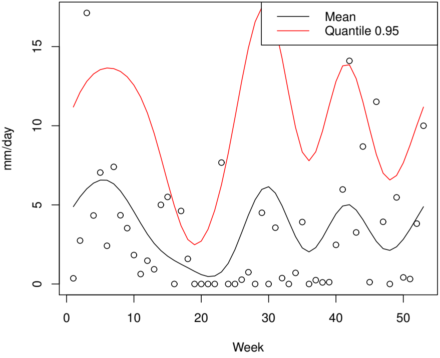


When is quantile regression useful

Some quantiles are more important than others:

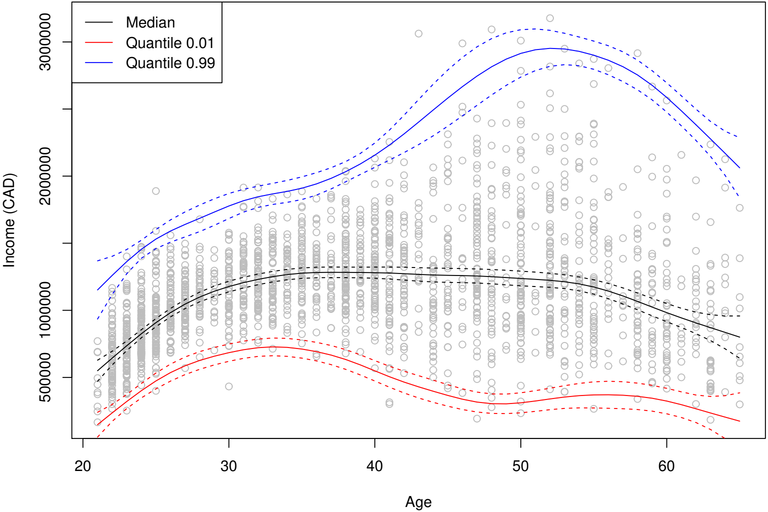
- electricity producers need to satisfy top electricity demand
- urban planners need estimates of extreme rainfall





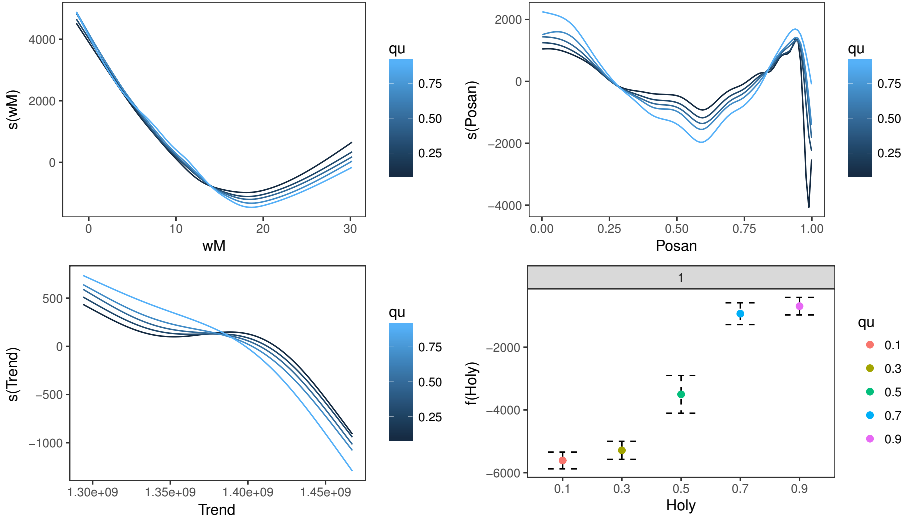
When is quantile regression useful

Effect of explanatory variables may depend on quantile



When is quantile regression useful

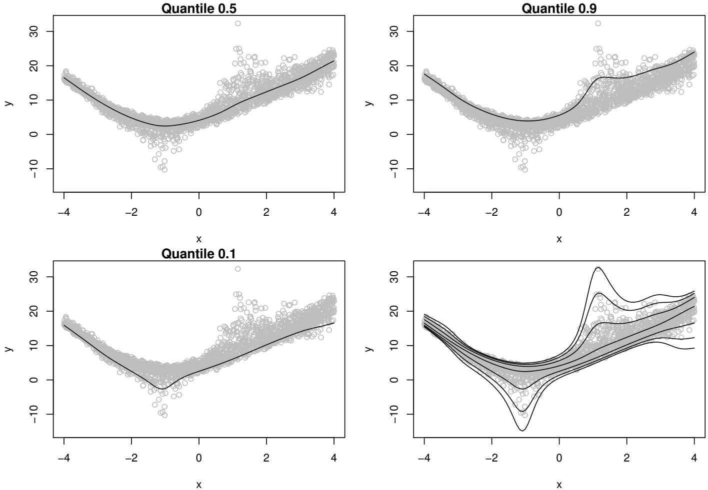
$$q_\tau(\text{Demand}) = f_1(\text{Temp}) + f_2(\text{TimeOfYear}) + f_3(\text{Trend}) + f_4(\text{Holiday}) + \cdots$$



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);



Beyond mean modelling: GAMLSS models

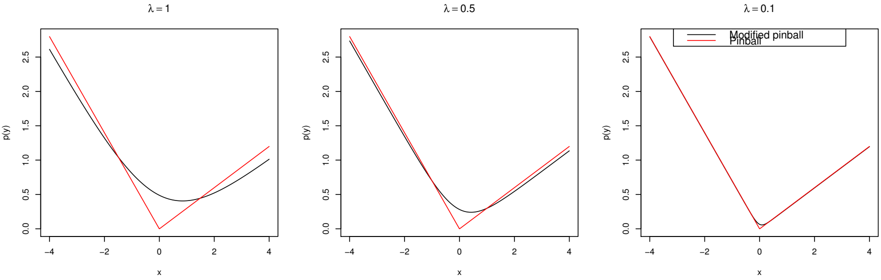
Structure:

1. Intro to GAMs for Location Scale and Shape
2. Intro to quantile GAMs
3. **GAMLSS and QGAM modelling in mgcv and 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 err $\in (0, 1)$ ($\downarrow$ 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, ...)
```

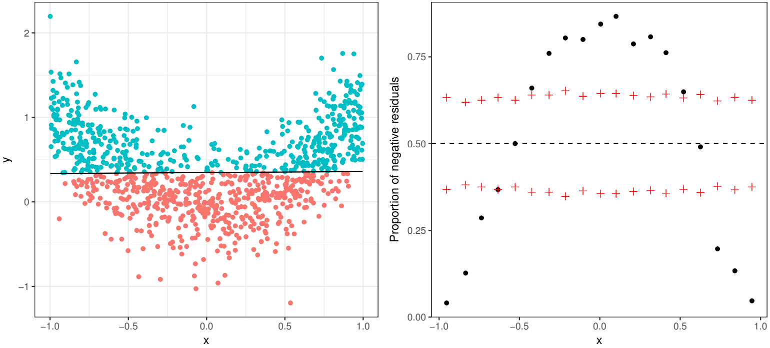
which is a compromise between bias and speed.

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)
```



Conclusions

THANK YOU!

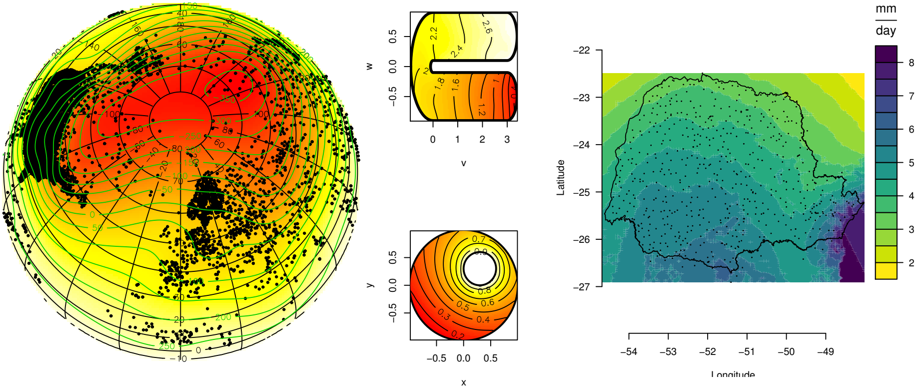


Figure: Examples of quantile GAMs from Fasiolo et al. (2017).

References I

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