Additive quantile regression 1

Matteo Fasiolo (University of Bristol, UK)

matteo.fasiolo@bristol.ac.uk

May 17, 2018

Today's plan

Morning session

1. What is quantile regression
2. When is it useful
3. Linear quantile mixed models
4. Non-parametric additive quantile models
5. Quantile regression using `qgam`
6. Hands-on session

Afternoon session

1. Fitting additive quantile models
2. Model checking
3. Types of smooths
4. More quantile regression using `qgam`
5. Hands-on session

Structure of the talk

Structure:

1. What is quantile regression
2. When is it useful
3. Linear quantile mixed models
4. Non-parametric additive quantile models
5. Quantile regression using `qgam`

Structure of the talk

Structure:

1. **What is quantile regression**
2. When is it useful
3. Linear quantile mixed models
4. Non-parametric additive quantile models
5. How to do 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

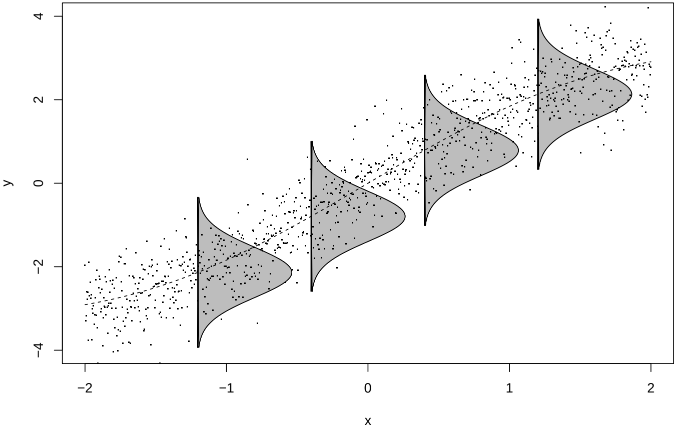


Figure : Gaussian model with variable mean.
In mgcv: `gam(y~s(x), family=gaussian)`.

What is quantile regression

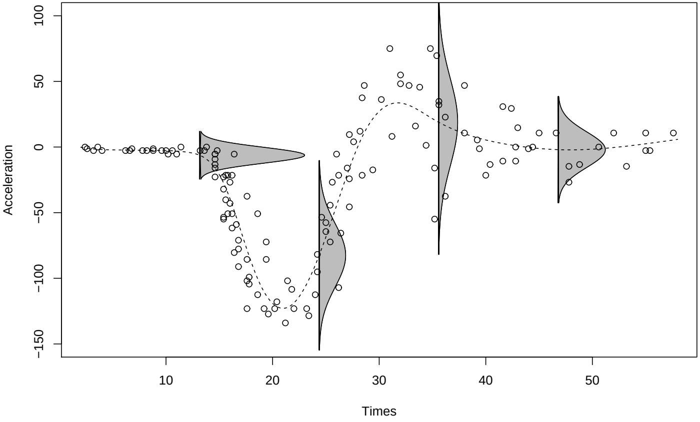


Figure : Gaussian model with variable mean and variance.
In mgcv: `gam(list(y~s(x), ~s(x)), family=gaulss)`.

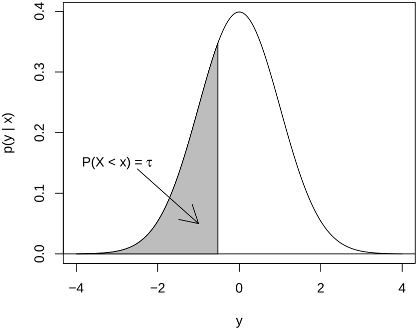
What is quantile regression

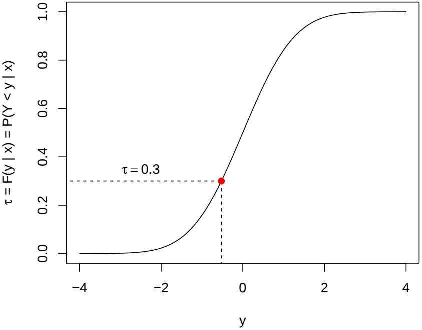
Lots of options for $p_m(y|\mathbf{x})$: binomial, gamma, Poisson, Tweedie...

We consider continuous (not discrete) y.

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

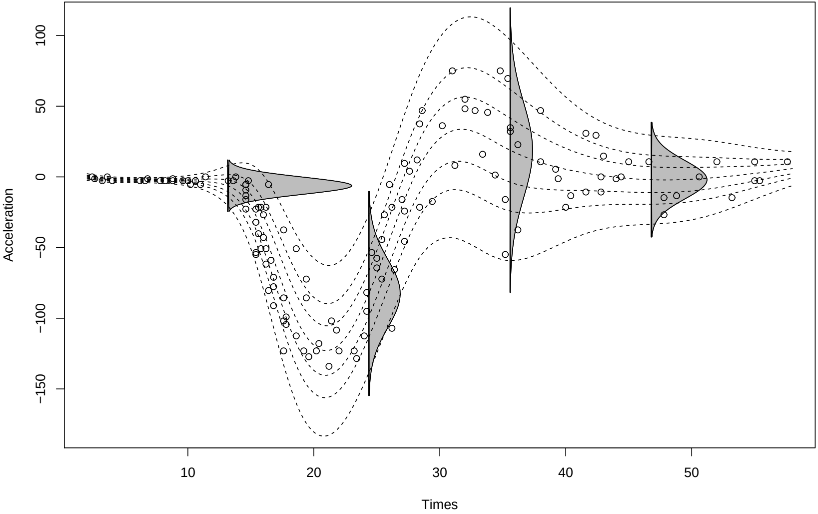
The τ-th ($\tau \in (0,1)$) quantile is $\mu_\tau(\mathbf{x}) = F^{-1}(\tau|\mathbf{x})$.





What is quantile regression

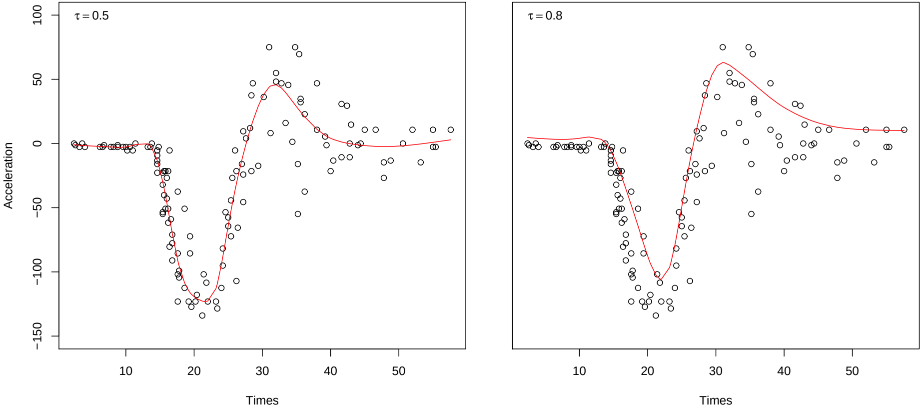
Given $p_m(y|\mathbf{x})$ we can get the conditional quantiles $\mu_\tau(\mathbf{x})$.



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

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

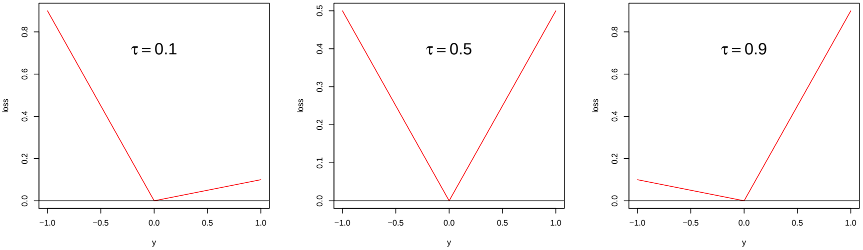
but also the minimizer of

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

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** $\mu_\tau(\mathbf{x}) = \boldsymbol{\beta}^\mathsf{T}\mathbf{x} = \beta_1 x_1 + \ldots \beta_p x_p$.

$\hat{\boldsymbol{\beta}}$ 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

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

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

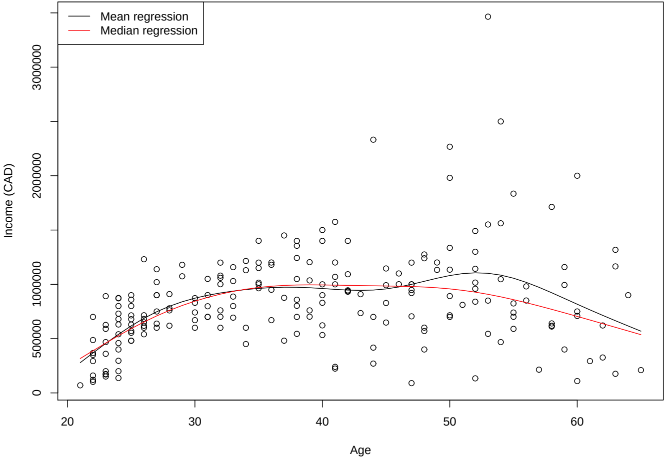
Structure of the talk

Structure:

1. What is quantile regression
2. **When is it useful**
3. Linear quantile mixed models
4. Non-parametric additive quantile models
5. How to do quantile regression using `qgam`

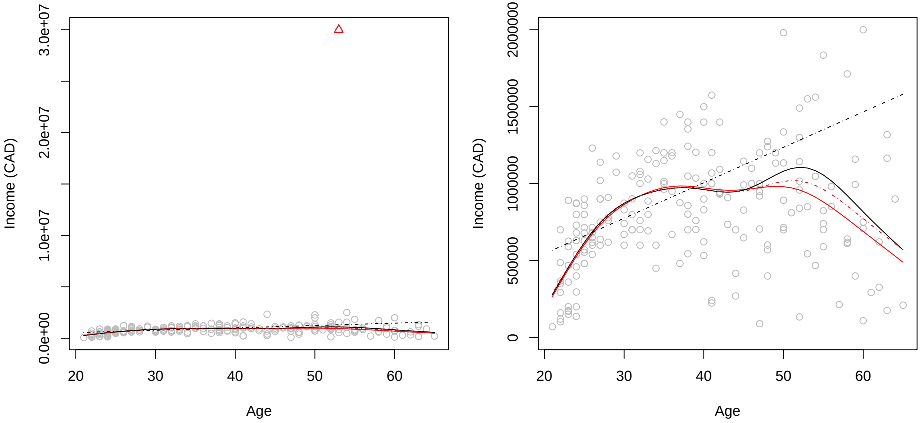
When is quantile regression useful

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



When is quantile regression useful

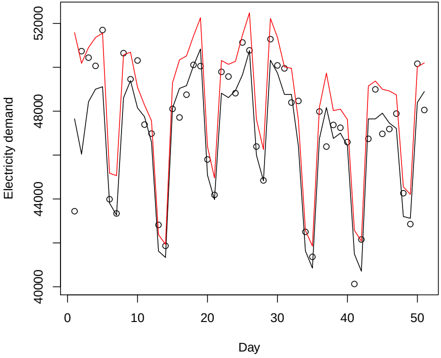
The median is also more **resistant to outliers**.

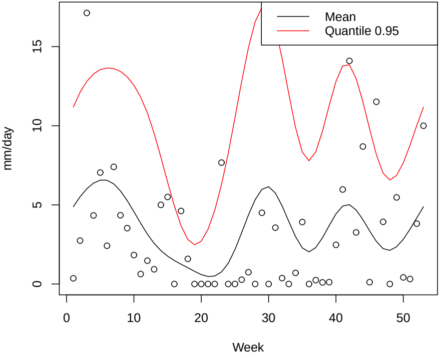


When is quantile regression useful

Some quantiles are more important than others:

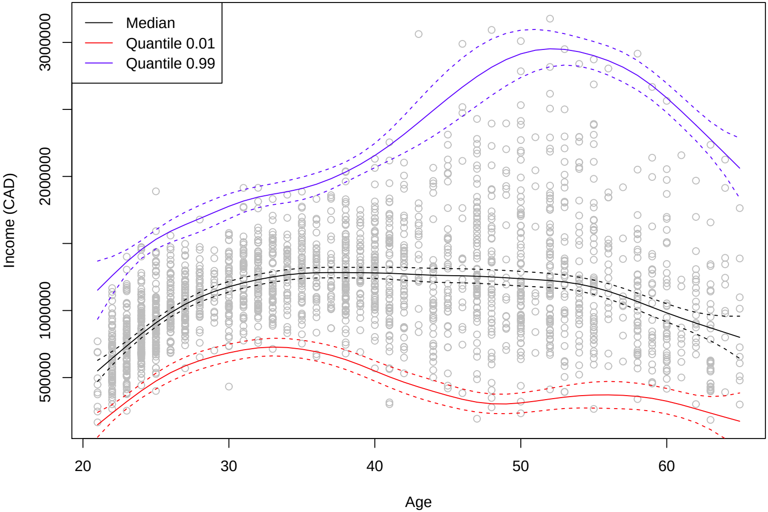
- electricity producers need to satisfy top electricity demand
- urban planners need estimates of extreme rainfall





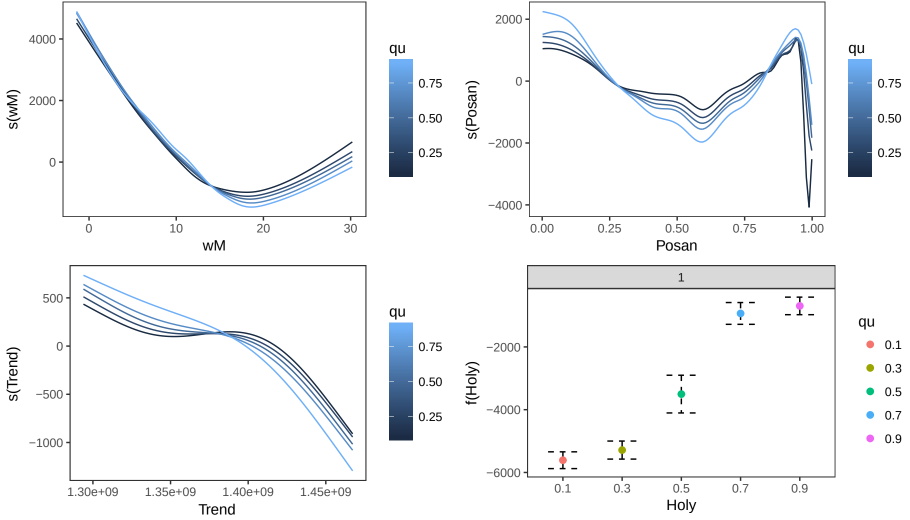
When is quantile regression useful

Effect of explanatory variables may depend on quantile



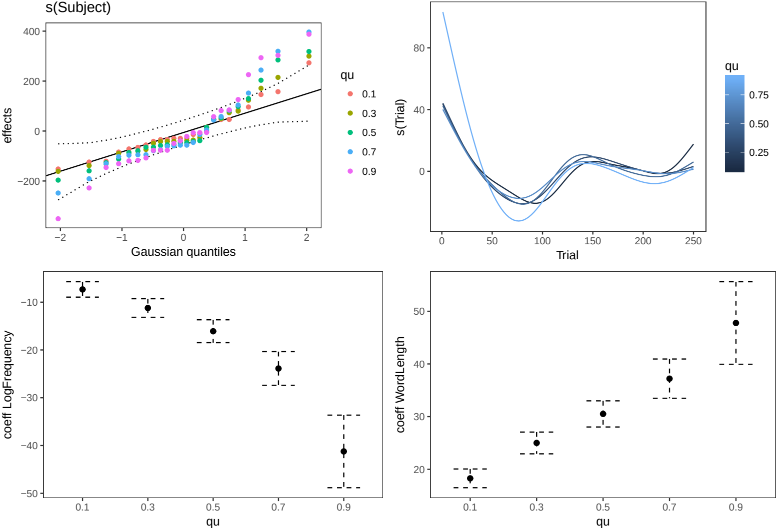
When is quantile regression useful

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



When is quantile regression useful

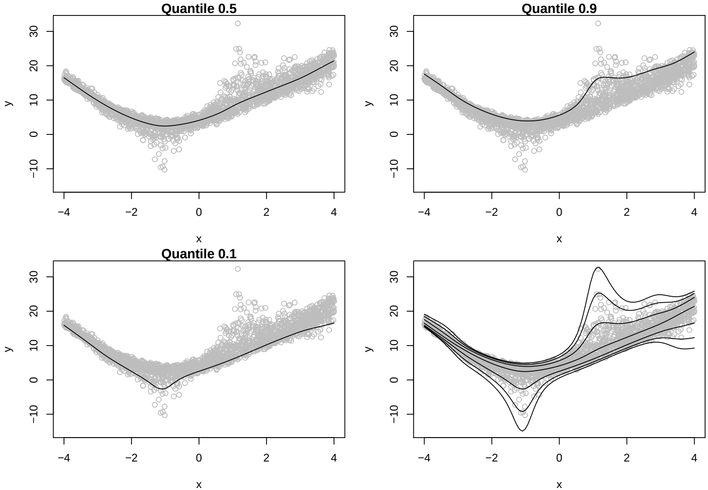
$$q_\tau(\mathrm{RT}) = f_1(\mathrm{Subject}) + f_2(\mathrm{Trial}) + f_3(\mathrm{WordFrequency}) + f_4(\mathrm{WordLength}) + \cdots$$



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. **Linear quantile mixed models**
4. Non-parametric additive quantile models
5. How to do quantile regression using `qgam`

Linear quantile mixed models

Suppose we have data on bone density (*bd*) as a function of *age*.

We have m subjects and n data pairs per subject

- subj 1: $\{bd_{11}, age_{11}\}, \dots, \{bd_{n1}, age_{n1}\}$
- subj j: $\{bd_{1j}, age_{1j}\}, \dots, \{bd_{nj}, age_{nj}\}$
- subj m: $\{bd_{1m}, age_{1m}\}, \dots, \{bd_{nm}, age_{nm}\}$

Standard linear quantile model ignores individual differences

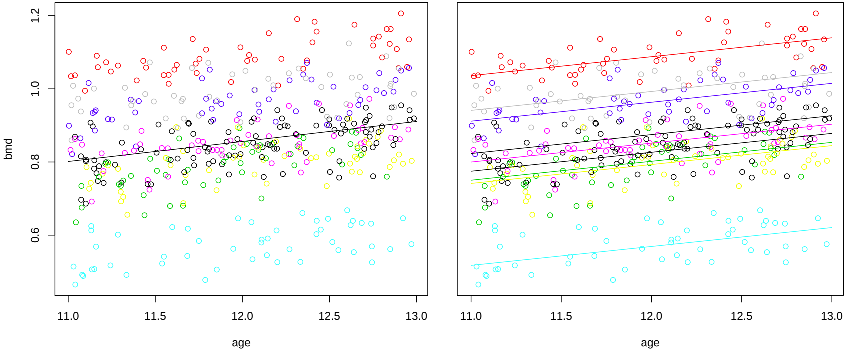
$$\mu_\tau(age_{ij}) = \alpha + \beta age_{ij}.$$

We can include random intercept per subject

$$\mu_\tau(age_{ij}) = \alpha + \beta age_{ij} + \mathsf{a}_j,$$

where $\mathbf{a} = \{\mathsf{a}_1, \dots, \mathsf{a}_m\} \sim N(\mathbf{0}, \mathbf{\Sigma})$.

Linear quantile mixed models



We can also include random slopes

$$\mu_\tau(age_{ij}) = \alpha + (\beta + b_j)age_{ij} + a_j,$$

where $\mathbf{a} \sim N(\mathbf{0}, \mathbf{\Sigma_a})$ and $\mathbf{b} \sim N(\mathbf{0}, \mathbf{\Sigma_b})$.

Linear quantile mixed models

In qgam (as in mgcv) random effect are specified as:

```
qgam(bmd ~ 1 + s(subject, bs = "re") +
           age + s(age, subject, bs = "re"), ...)
```

In simplest case $\boldsymbol{\Sigma}_{\mathbf{a}} = \gamma_{\mathbf{a}} \mathbf{I}$ and $\boldsymbol{\Sigma}_{\mathbf{b}} = \gamma_{\mathbf{b}} \mathbf{I}$, that is

$$\boldsymbol{\Sigma}_{\mathbf{a}} = \begin{bmatrix} \gamma_{\mathbf{a}} & 0 & 0 & \dots & 0 \\ 0 & \gamma_{\mathbf{a}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \gamma_{\mathbf{a}} \end{bmatrix}$$

Variances $\gamma_{\mathbf{a}}$ and $\gamma_{\mathbf{b}}$ must be estimated (afternoon session).

Structure of the talk

Structure:

1. What is quantile regression
2. When is it useful
3. Linear quantile mixed models
4. **Non-parametric additive quantile models**
5. How to do quantile regression using `qgam`

Non-parametric additive quantile models

In additive modelling

$$\mu_\tau(\mathbf{x}) = \sum_{j=1}^{m} f_j(\mathbf{x}),$$

where f_j can be fixed, random or smooth effects. Example

$$\mu_\tau(age_{ij}) = \alpha + \mathrm{a}_j + f(age_{ij})$$

where f a non-linear smooth function.

Smooth effects built using spine bases

$$f(age) = \sum_{k=1}^{r} \beta_k b_k(age)$$

where β_k are unknown coeff and $b_k(age)$ are known spline basis functions.

Non-parametric additive quantile models

Example 1: B-splines

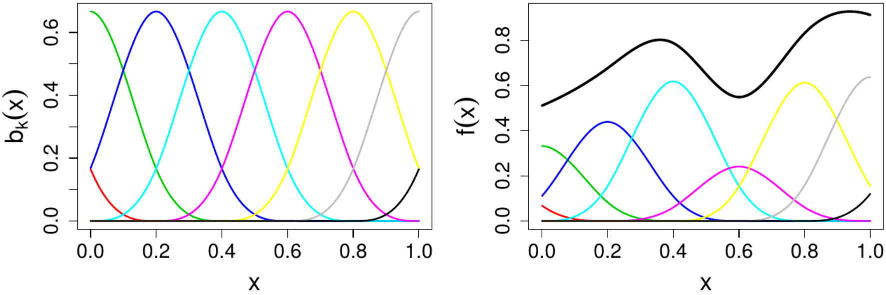


Figure : B-spline basis (left) and smooth (right). Courtesy of Simon Wood.

Non-parametric additive quantile models

Example 2: Thin plate regression splines (TPRS)

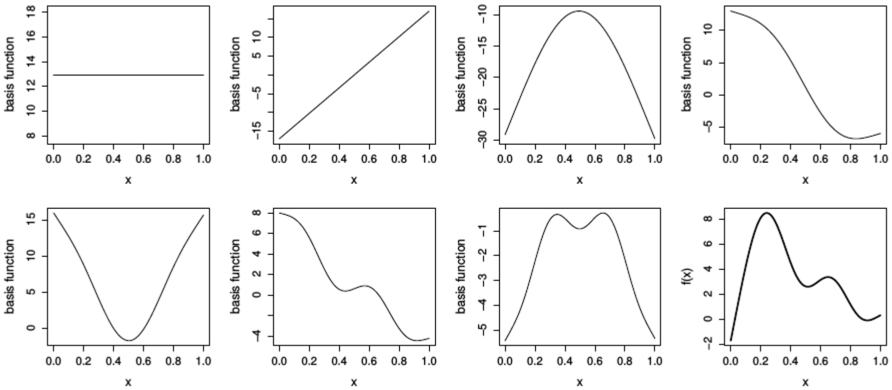


Figure : Rank 7 TPRS basis. Image from Wood (2006).

Non-parametric additive quantile models

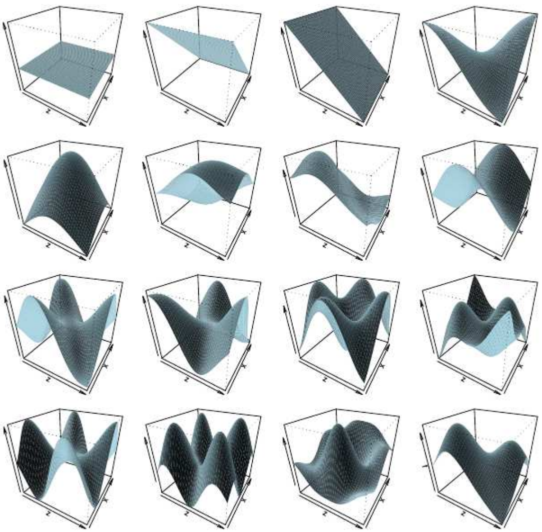


Figure : Rank 17 2D TPRS basis. Courtesy of Simon Wood.

Non-parametric additive quantile models

In general

$$f(\mathbf{x}) = \sum_{k=1}^{r} \beta_k b_k(\mathbf{x}).$$

To determine complexity of $f(\mathbf{x})$:

- the basis rank r is large enough for sufficient flexibility
- a Gaussian prior on $\boldsymbol{\beta}$ controls the wiggliness of the effects

In morning practical we'll see only 1D effects.

More complex effects explained in the afternoon.

In `qgam` or `mgcv`:

```
qgam(y ~ 1 + s(x1, bs = "tp") + s(x2, bs = "cr"), ...)
```

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.

Structure of the talk

Structure:

1. What is quantile regression
2. When is it useful
3. Linear quantile mixed models
4. Non-parametric additive quantile models
5. **How to do quantile regression using qgam**

THANK YOU!

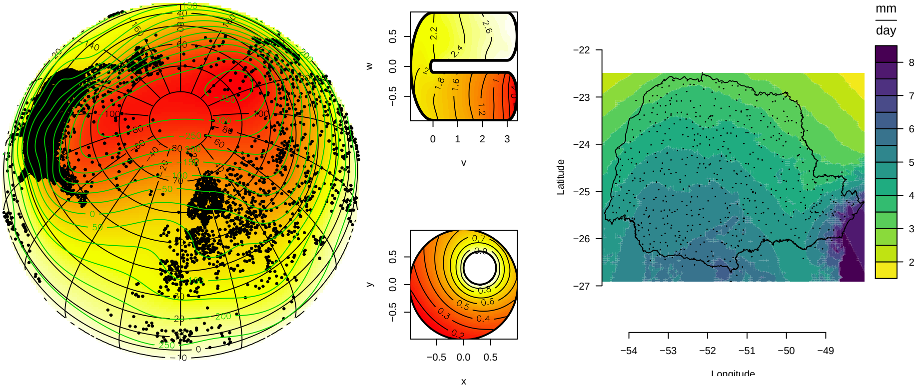


Figure : Examples of quantile GAMs from Fasiolo et al. (2017).

References I

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