Additive quantile regression 2

Matteo Fasiolo (University of Bristol, UK)

matteo.fasiolo@bristol.ac.uk

May 17, 2018

Structure of the talk

Structure:

1. Fitting additive quantile models
2. Convergence checking
3. Types of smooths
4. More quantile regression using `qgam`

Structure of the talk

Structure:

1. **Fitting additive quantile models**
2. Convergence checking
3. Types of smooths
4. More quantile regression using `qgam`

Fitting additive quantile models

Here I describe our approach to estimating quantile GAMs.

This is implemented in `qgam` and in `mgcv`.

I explain the main ideas, for details see Fasiolo et al. (2017).

Recall $\mu_\tau(\mathbf{x}) = \sum_{j=1}^m f_j(\mathbf{x})$ where $f_j(\mathbf{x})$ can be

- parametric e.g. $f_j(\mathbf{x}) = \alpha x_j + \beta x_j^2$
- random effects
- spline-based smooths such as

$$f_j(x_j) = \sum_{i=1}^{r} \beta_{ji} b_{ji}(x_j)$$

where β_{ji} are coefficients and $b_{ji}(x_j)$ are known spline basis functions.

Fitting additive quantile models

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

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

When random or smooth effects included, minimized penalized loss

$$\hat{\boldsymbol{\beta}} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{\sigma} L_y(\boldsymbol{\beta}) + \text{Pen}(\boldsymbol{\beta}|\boldsymbol{\gamma}) \right\},$$

where

- $L_y(\boldsymbol{\beta})$ quantifies goodness of fit
- $\text{Pen}(\boldsymbol{\beta}|\boldsymbol{\gamma})$ penalizes complexity
- $\boldsymbol{\gamma}$ smoothing parameters ($\uparrow \boldsymbol{\gamma}$ $\uparrow$smoothness)
- $1/\sigma$ is learning rate

Fitting additive quantile models

We use a hierarchical fitting framework:

1. Calibrate $1/\sigma$ to have correct coverage of confidence intervals

2. For fixed $1/\sigma$:
   select γ to have "right" amount of smoothness

3. For fixed $1/\sigma$ and γ:
   estimate $\boldsymbol{\beta}$ to determine actual fit

Fitting additive quantile models

We use a hierarchical fitting framework:

1. Calibrate $1/\sigma$ to have correct coverage of confidence intervals

$$\hat{\sigma} = \underset{\sigma}{\text{argmin}} \; \text{IKL}(\sigma).$$

2. For fixed $1/\sigma$:
   select $\boldsymbol{\gamma}$ to have "right" amount of smoothness

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}}{\text{argmin}} \; \hat{G}(\boldsymbol{\gamma}|\sigma).$$

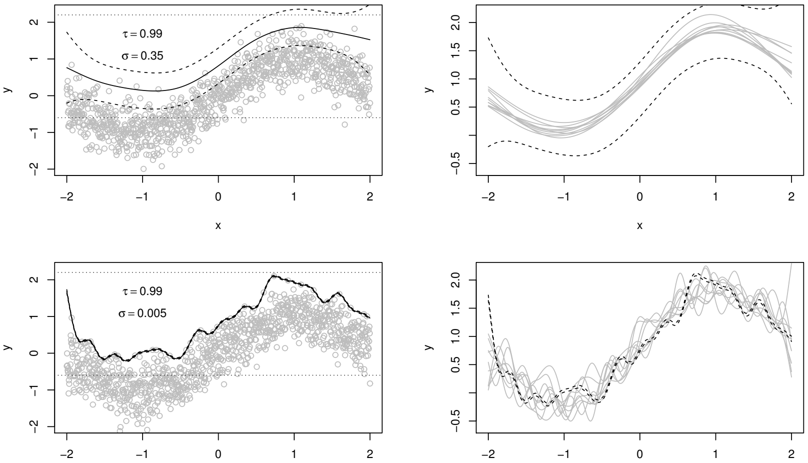
3. For fixed $1/\sigma$ and $\boldsymbol{\gamma}$:
   estimate $\boldsymbol{\beta}$ to determine actual fit

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \; V(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma).$$

Fitting additive quantile models

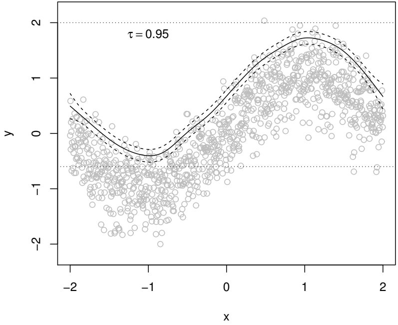
Effect of σ and γ is interrelated:

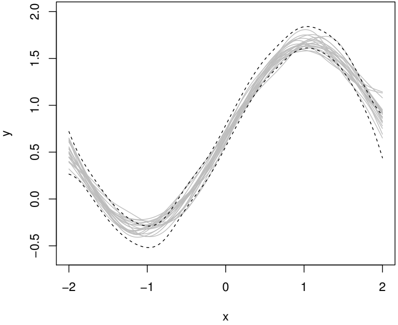
- credible intervals: $\uparrow \sigma \rightarrow \uparrow$ intervals width
- complexity of fit: $\uparrow \sigma \rightarrow \uparrow \hat{\gamma} \rightarrow \uparrow$ smoothness



Fitting additive quantile models

We choose σ so that confidence intervals for $\mu_\tau(\mathbf{x})$ are well calibrated.



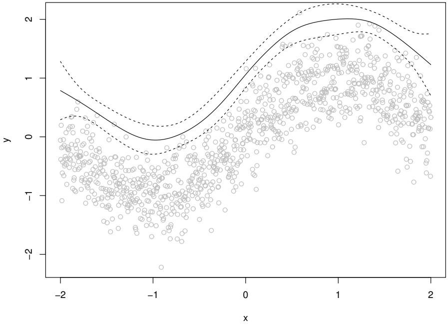


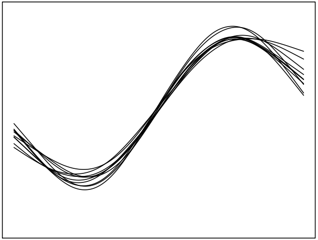
This results also in good smoothness selection for $\gamma|\sigma$.

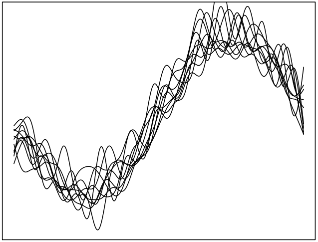
Fitting additive quantile models

The hierarchical fitting framework:

1. Calibrate $1/\sigma$
2. Select $\boldsymbol{\gamma}|\sigma$
3. Estimate $\boldsymbol{\beta}|\boldsymbol{\gamma},\sigma$ to determine fit





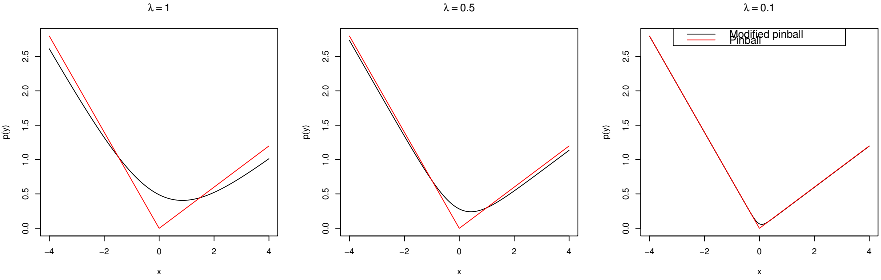


Fitting additive quantile models

We use a modified loss based on the Extended log-F (ELF) density:

$$\tilde{\rho}(y|\mu, \sigma, \tau, \lambda) \propto -(1-\tau)\frac{y-\mu}{\sigma} + \lambda \log\left[1 + e^{\frac{y-\mu}{\lambda\sigma}}\right],$$

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



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

Fitting additive quantile models

Increasing `err` leads to:

- faster and more stable computation
- more bias

Interpretation, if μ^* is minimizer of ELF loss:

$$|F(\mu^*) - \tau| \leq \text{err}$$

`err` is an upper bound on the bias.

By default:

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

Structure of the talk

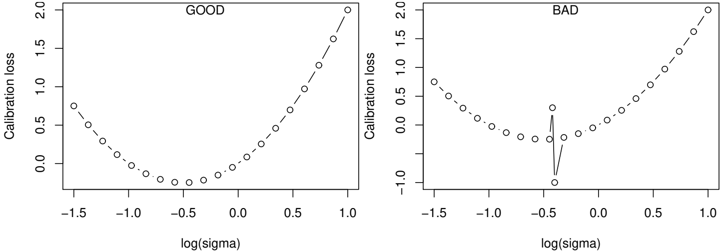
Structure:

1. Fitting additive quantile models
2. **Convergence checking**
3. Types of smooths
4. More quantile regression using `qgam`

Convergence checking

There might be problems at each step:

1. Calibrate $1/\sigma$



2. Select $\boldsymbol{\gamma}|\sigma$

```
Warning: outer Newton did not converge fully
```

3. Estimate $\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma$ to determine fit

```
Error: In gam.fit5(...): iteration limit reached
```

Convergence checking

Problems might be due to structural problems:

```
fit <- qgam(y <- x1 + s(x1), qu = 0.5)

Error in chol.default(iS) :
  the leading minor of order 11 is not positive definite
```

In general increasing `err` alleviates problem but increases bias.

`mgcViz` and `qgam` provide some diagnostics.

Structure of the talk

Structure:

1. Fitting additive quantile models
2. Convergence checking
3. **Types of smooths**
4. More quantile regression using `qgam`

Types of smooths

`mgcv` offers a wide variety of smooths (see `?smooth.terms`).

Univariate types:

- `s(x) = s(x, bs = "tp")` thin-plate-splines
- `s(x, bs = "cr")` cubic regression spline
- `s(x, bs = "ad")` adaptive smooth

Multivariate type:

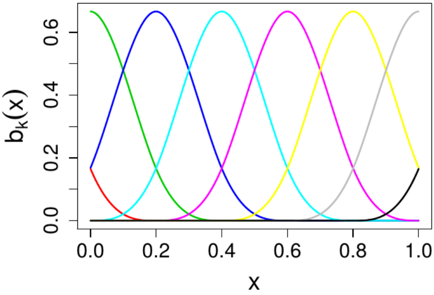
- `s(x1, x2) = s(x1, x2, bs = "tp")` thin-plate-splines (isotropic)
- `te(x1, x2)` tensor-product-smooth (anisotropic)
- `s(x, y, bs = "sos")` smooth on sphere

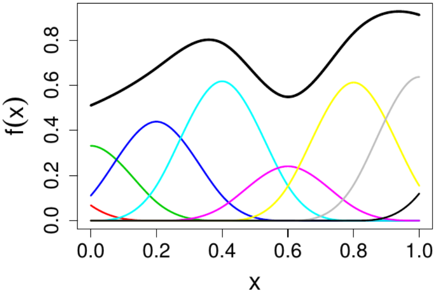
They can depends on factors:

- `s(x, by = Subject)`
- `s(x, Subject, bs = "fs")`

Types of smooths

`s(x, bs = "cr", k = 20)`



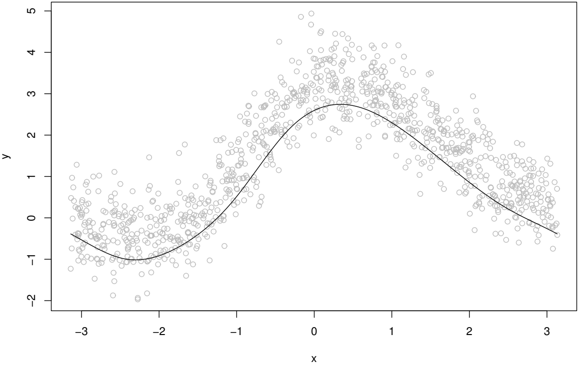


Cubic regression spline are related to the optimal solution to

$$\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \gamma \int f''(x)^2 \, dx.$$

Types of smooths

`s(x, bs = "cc")`

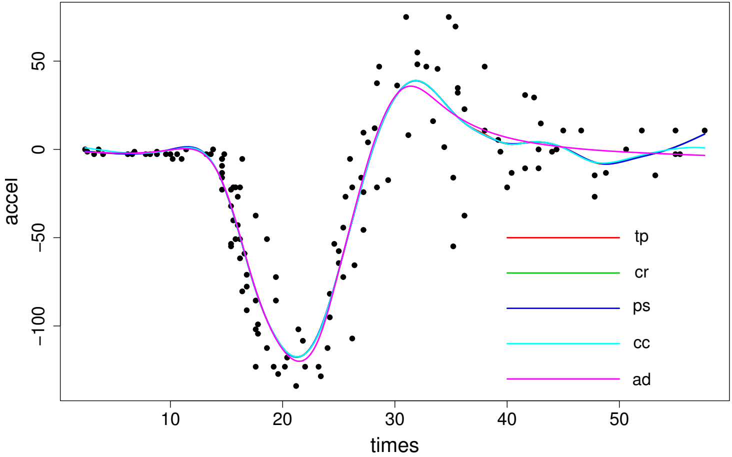


Cyclic cubic regression spline make so that

- $f(x_{min}) = f(x_{max})$
- $f'(x_{min}) = f'(x_{max})$

Types of smooths

`s(x, bs = "ad")`



The wiggliness or smoothness of $f(x)$ depends on x.

Types of smooths

`s(x1, x2)`, `s(x1, x2, x3)`, ...

Based on thin plate regression splines basis.

Related to optimal solution to:

$$\sum_i \{y_i - f(x_i, z_i)\}^2 + \gamma \int f_{xx}^2 + 2f_{xz}^2 + f_{zz}^2 dxdz$$

A single smoothing parameter γ.

Isotropic: same smoothness along x_1, x_2, ...

Types of smooths

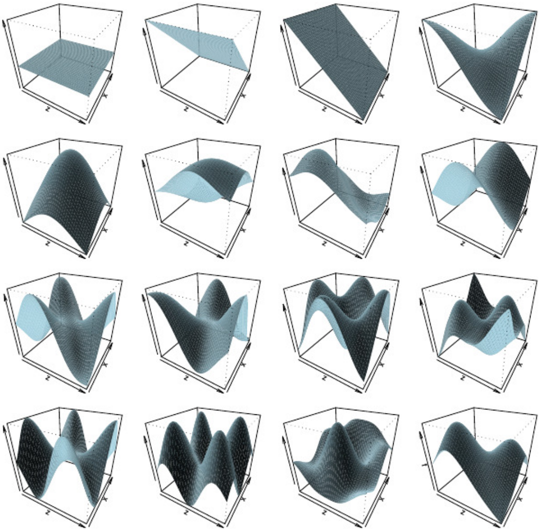


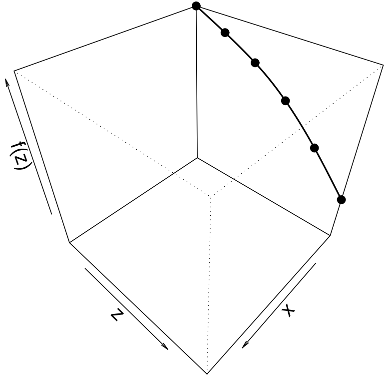
Figure: Rank 17 2D TPRS basis. Courtesy of Simon Wood.

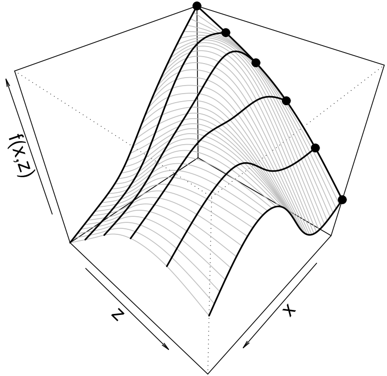
Types of smooths

Isotropic effect of x_1, x_2 are in same unit (e.g. Km).

If different units better use tensor product smooths `te(x1, x2)`.

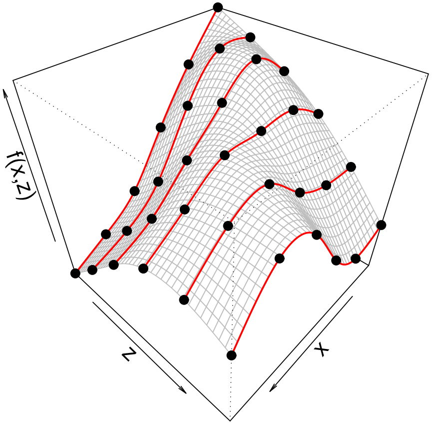
Construction: make a spline $f_z(z)$ a function of x by letting its coefficients vary smoothly with x

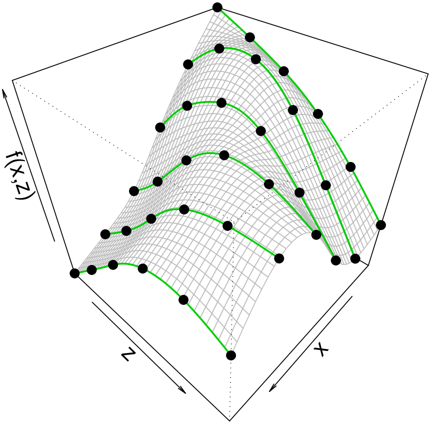




Types of smooths

- x-penalty: average wiggliness of red curves
- z-penalty: average wiggliness of green curves





Types of smooths

Can use (almost) any kind of marginal:

- `te(x1, x2, x2)` product of 3 thin-plate-spline bases
- `te(x1, x2, bs = c("cc", "cr"), k = c(10, 6))`
- `te(LO, LA, t, d=c(2,1), k=c(20,10), bs=c("tp","cc"))`

Basis of `te` contains functions of the form $f(x_1) + f(x_2)$.

To fit $f(x_1) + f(x_2) + f(x_1, x_2)$ separately use:

```
y ~ ti(x1) + ti(x2) + ti(x1, x2)
```

Types of smooths

By-factor smooths

Approach (1) is `s(x, by = subject)`, which means

- $\rho_\tau(x) = f_1(x) + ...$ if subject $= 1$
- $\rho_\tau(x) = f_2(x) + ...$ if subject $= 2$
- ...

Approach (2) is `s(x, subject, bs = "fs")`, which means

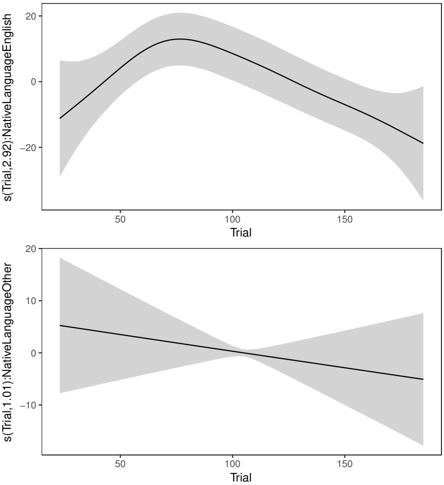
- $\rho_\tau(x) = b_1 + f_1(x) + ...$ if subject $= 1$
- $\rho_\tau(x) = b_2 + f_2(x) + ...$ if subject $= 2$
- ...

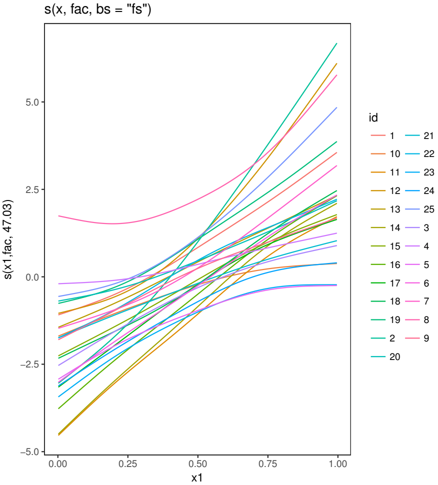
where $b_1, b_2, \cdots \sim N(0, \gamma_{\mathbf{b}}\mathbf{I})$ are random effects.

In (1) each f_j has its own smoothing parameter.

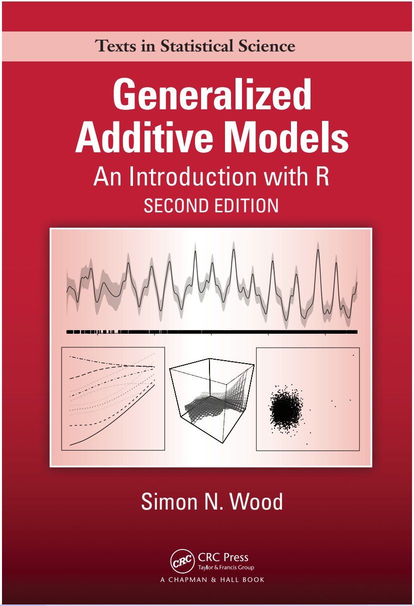
In (2) all f_j's have the same smoothing parameter.

Types of smooths





Further reading



Structure of the talk

Structure:

1. Fitting additive quantile models
2. Convergence checking
3. Types of smooths
4. **More quantile regression using qgam**

References I

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