Quantile regression coefficients modelling

Paolo Frumento

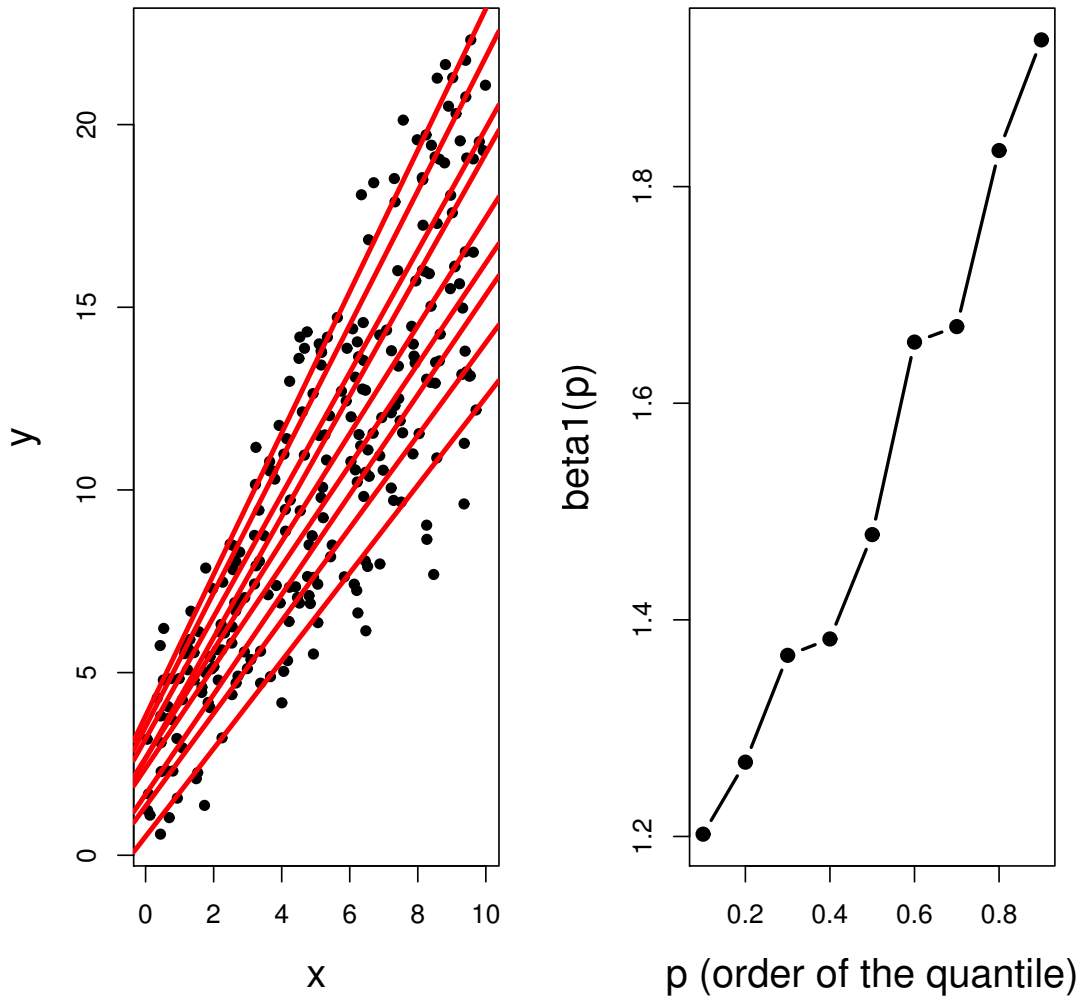
University of Pisa

Other participants

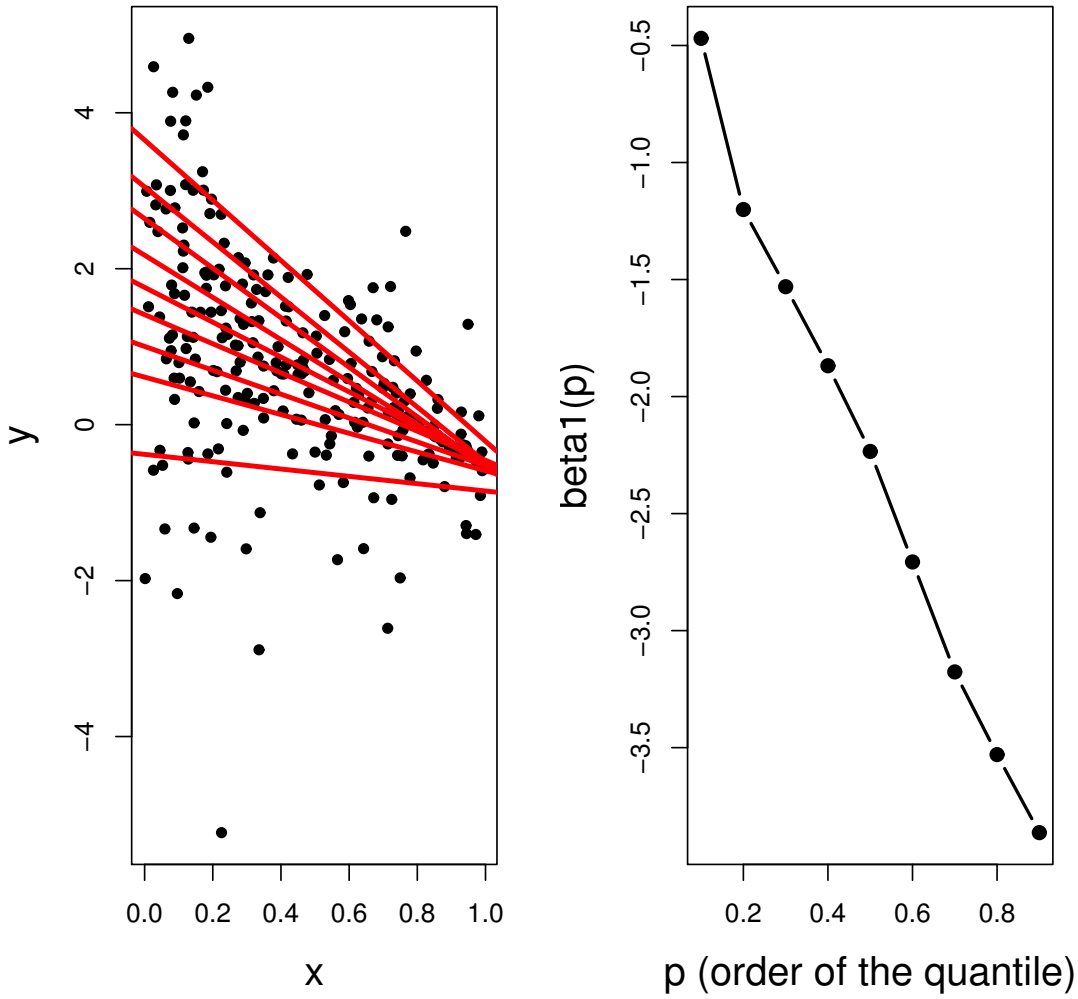
- Gianluca Sottile (University of Palermo, Italy)
- Ivan Fernández-Val (Boston University, US)
- Nicola Salvati (University of Pisa, Italy)
- Marcello Chiodi (University of Palermo, Italy)
- Matteo Bottai (Karolinska Institute, Stockholm, Sweden)

Quantile regression

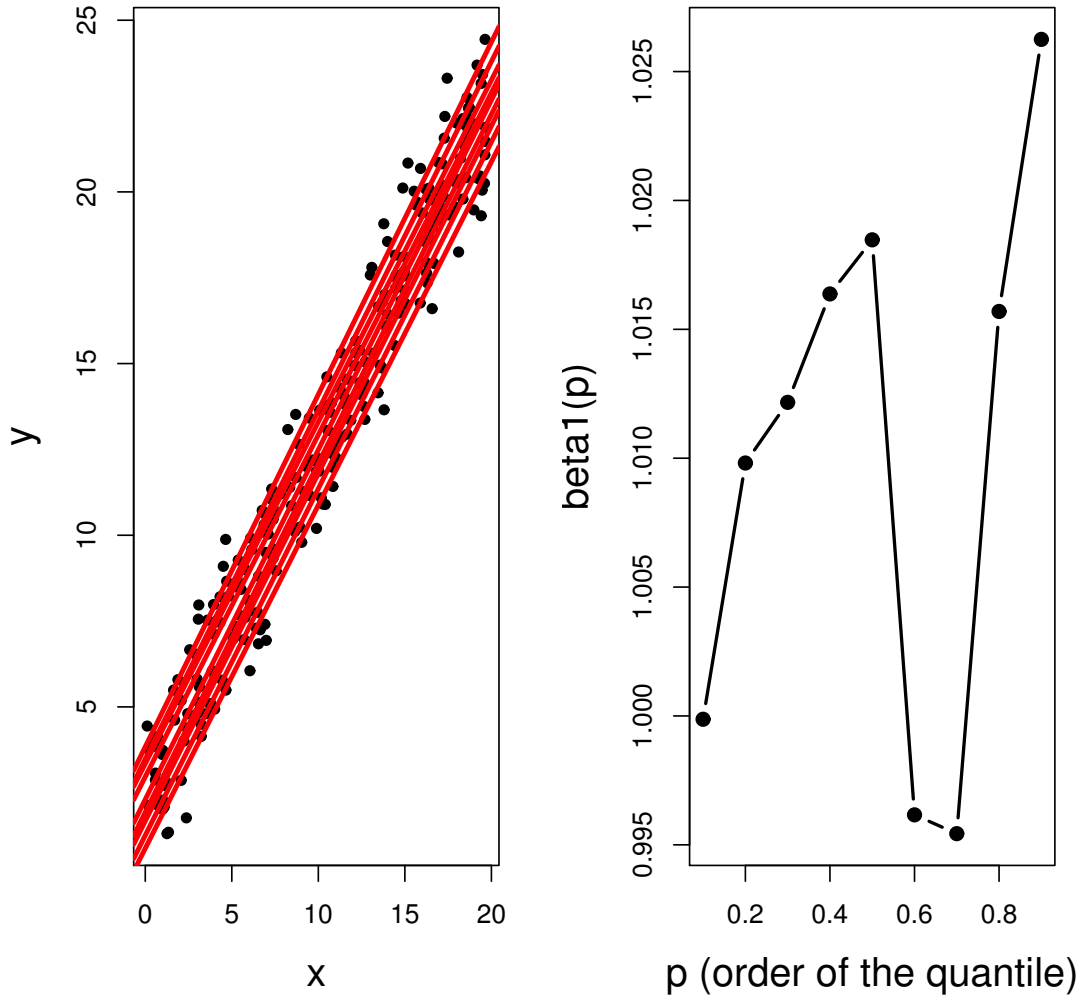
$$Q(p \mid x) = \beta_0(p) + \beta_1(p)x$$



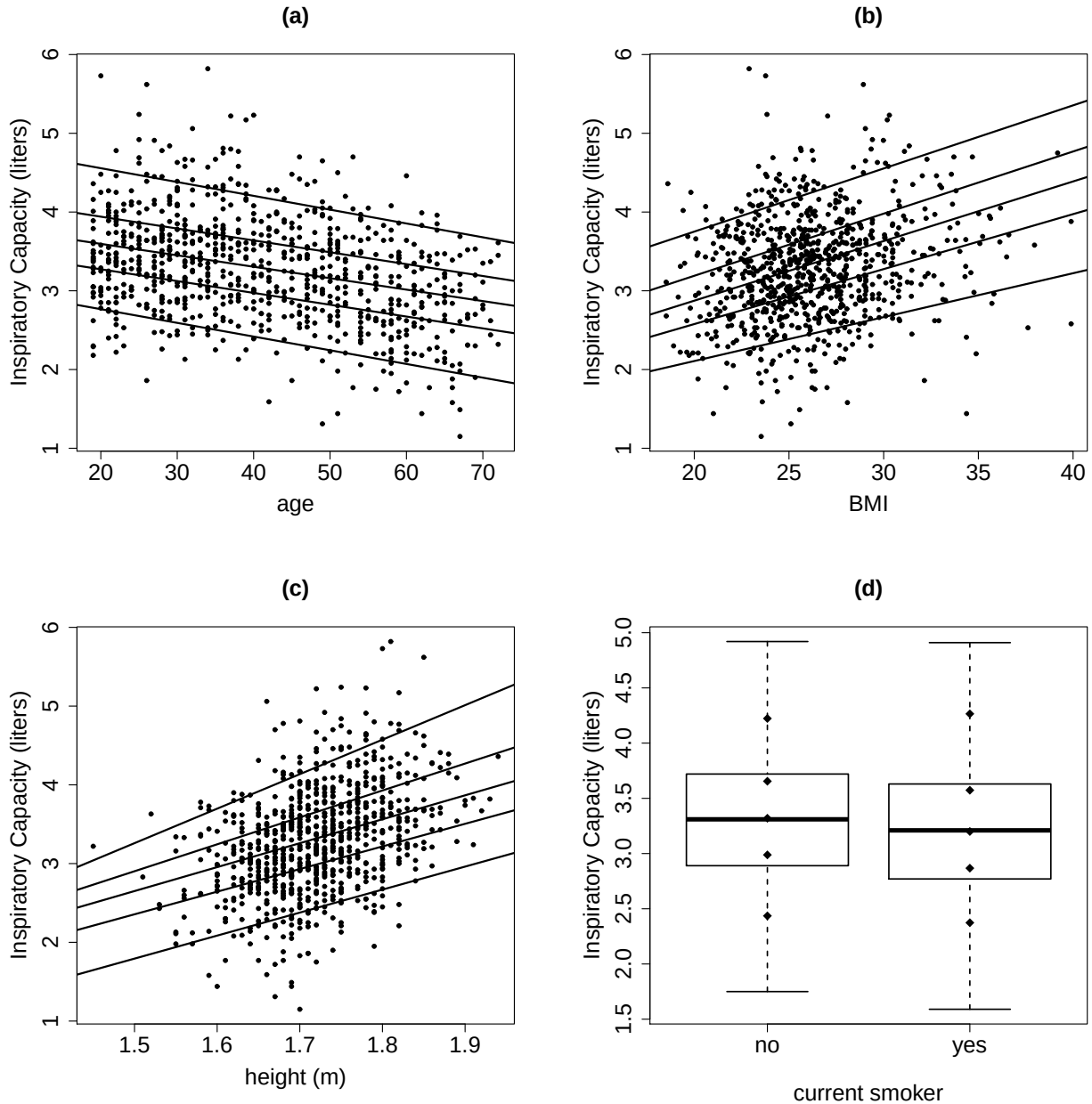
Quantile regression (2)



Quantile regression (3)

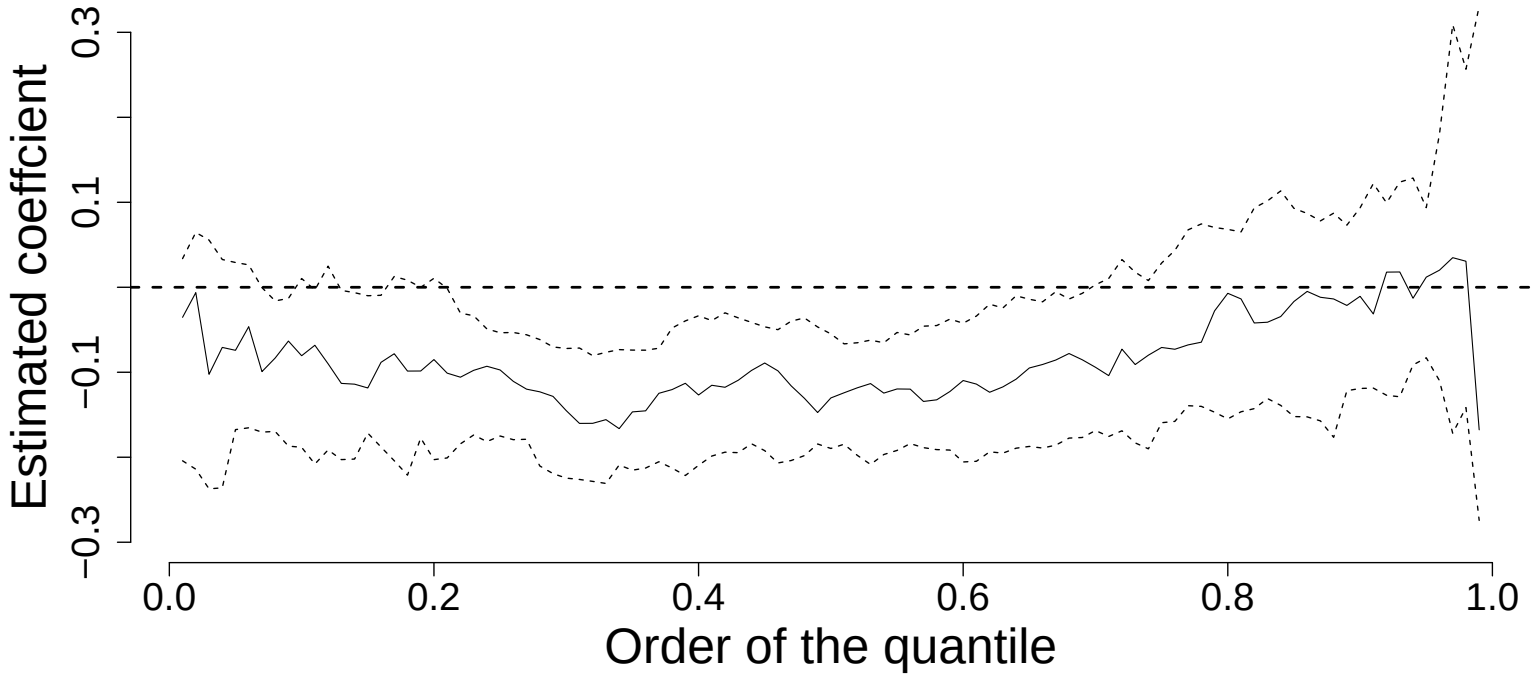


Example - Predictors of inspiratory capacity



Example (cont.)

Coefficient function of smoking, $p = (0.01, 0.02, \ldots, 0.99)$.

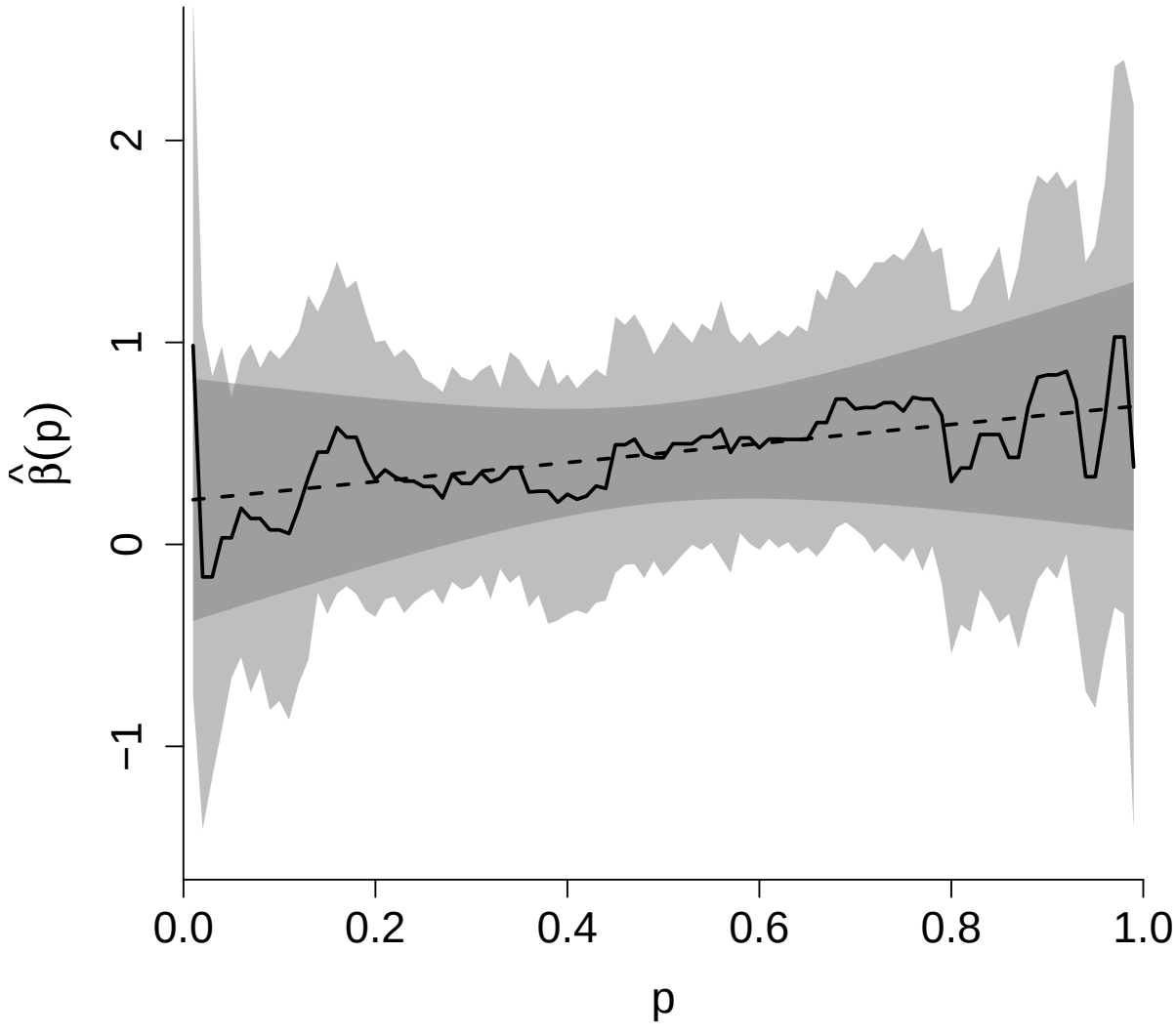


- Difficult to interpret
- Inefficient
- Instinctive visual interpolation

Possible solutions:

- Smoothing
- Modelling!

Quantile regression coefficients modelling (QRCM)



Quantile regression coefficients modelling (QRCM)

Linear effect of covariates on the quantile function:

$$Q(p \mid \boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\beta}(p)$$

A parametric model for $\boldsymbol{\beta}(p) = \{\beta_1(p), \ldots, \beta_q(p)\}$:

$$\beta_j(p \mid \boldsymbol{\theta}) = \theta_{j1} b_1(p) + \ldots + \theta_{jk} b_k(p)$$

In matrix form:

$$\boldsymbol{\beta}(p \mid \boldsymbol{\theta}) = \boldsymbol{\theta} \boldsymbol{b}(p)$$

where

$$\boldsymbol{b}(p) = [b_1(p), \ldots, b_k(p)]^{\mathrm{T}}$$

and $\boldsymbol{\theta}$ is a $q \times k$ matrix.

A parametric quantile function

$$Q(p \mid \boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{x}^T \beta(p \mid \boldsymbol{\theta}) = \boldsymbol{x}^T \boldsymbol{\theta} \boldsymbol{b}(p)$$

How to define $\boldsymbol{b}(p)$?

Examples

$$Q(p \mid x, \boldsymbol{\theta}) = \beta_0(p \mid \boldsymbol{\theta}) + \beta_1(p \mid \boldsymbol{\theta})x$$

Example (1)

$$\beta_0(p) = \theta_{00} + \theta_{01} p$$

$$\beta_1(p) = \theta_{10} + \theta_{11} p$$

Linear functions of p.

- $\theta_0 + \theta_1 p$ is the quantile function of a $U(\theta_0, \theta_0 + \theta_1)$.
- A new interpretation!
- $Q(p \mid x, \boldsymbol{\theta})$ well defined if $\theta_{01} + \theta_{11} x > 0$ for all x.
- $\theta_{11} = 0$ forces homoskedasticity.
- If $\theta_{00} = \theta_{01} = 0$ (no intercept), a zero-inflated model.

Example (1) - cont.

$$\boldsymbol{b}(p) = \begin{pmatrix} 1 \\ p \end{pmatrix} \text{ and } \boldsymbol{\theta} = \begin{pmatrix} \theta_{00} & \theta_{01} \\ \theta_{10} & \theta_{11} \end{pmatrix}$$

Example (2)

$$\beta_0(p) = \theta_{00} + \theta_{01} z(p)$$

$$\beta_1(p) = \theta_{10} + \theta_{12} p,$$

with $z(p)$ the quantile function of a standard Normal.

- A "mix" between Uniform and Normal
- No closed form PDF!
  - This distribution is only defined through its quantile function.
  - If $\theta_{12} = 0$, we have the standard linear model with coefficients $\beta_0 = \theta_{00}$ and $\beta_1 = \theta_{10}$, and standard deviation $\sigma = \theta_{01}$.

Example (2) - cont.

$$\boldsymbol{b}(p) = \begin{pmatrix} 1 \\ z(p) \\ p \end{pmatrix} \text{ and } \boldsymbol{\theta} = \begin{pmatrix} \theta_{00} & \theta_{01} & 0 \\ \theta_{10} & 0 & \theta_{12} \end{pmatrix}$$

Example (3)

$$\beta_0(p \mid \boldsymbol{\theta}) = \theta_{00} + \theta_{01} p + \theta_{02} p^2$$

$$\beta_1(p \mid \boldsymbol{\theta}) = \theta_{10} + \theta_{11} \log(p) + \theta_{12} \cos(p)$$

- $\boldsymbol{b}(p)$ must induce a well-defined QF for some $\boldsymbol{\theta}$
- Use meaningful assumptions (boundedness, positivity)
- $\boldsymbol{b}(p)$ can have asymptotes

Example (4)

$$\beta_0(p \mid \boldsymbol{\theta}) = \theta_{00} + \theta_{01} \log(p) + \theta_{02} \log(1 - p)$$

$$\beta_1(p \mid \boldsymbol{\theta}) = \theta_{10} + \theta_{13} p$$

- $\beta_0(p)$ unbounded
- $\beta_1(p)$ monotone, bounded between θ_{10} (when $p = 0$) and $\theta_{10} + \theta_{13}$ (when $p = 1$)
- Special cases: Uniform, asymmetric Logistic, Logistic, (shifted) Exponential

The estimator

Ordinary quantile regression for the pth quantile: minimise

$$L(\beta(p)) = \sum_{i=1}^{n} (p - \omega_i(p))(y_i - \boldsymbol{x}_i^{\mathrm{T}}\beta(p))$$

where $\omega_i(p) = I(y_i \leq \boldsymbol{x}_i^{\mathrm{T}}\beta(p))$.

Our estimator: minimise

$$\overline{L}(\boldsymbol{\theta}) = \int_0^1 L(\beta(p \mid \boldsymbol{\theta}))\mathrm{d}p.$$

- Average loss function
- Estimating "all" quantiles at once
- Not a likelihood

The gradient function

Ordinary QR: find the approximated zeroes of

$$S(\beta(p)) = \sum_{i=1}^{n} \boldsymbol{x}_i(\omega_i(p) - p).$$

Our estimator: find the zeroes of

$$\overline{S}(\boldsymbol{\theta}) = \int_0^1 S(\beta(p \mid \boldsymbol{\theta}))\boldsymbol{b}(p)^{\mathrm{T}}\mathrm{d}p.$$

Note

$\overline{L}(\boldsymbol{\theta})$ and $\overline{S}(\boldsymbol{\theta})$ can be written in a closed form (well, almost)

Properties

- The entire quantile function is modelled at once
- Smooth loss function (simple computation and asymptotics)
- No need of using bootstrap or estimating the sparsity function
- You can take the integral over (p_1, p_2) instead of $(0, 1)$...
- ... For standard QR, use $p_1 = p_2 = p$...
- ... Or, use a very flexible parametrisation
- More parsimonious and efficient than QR
- $\overline{S}(\boldsymbol{\theta}) = 0$ if the fitted CDF values are "as uniform as possible"

Predictors of inspiratory capacity (cont.)

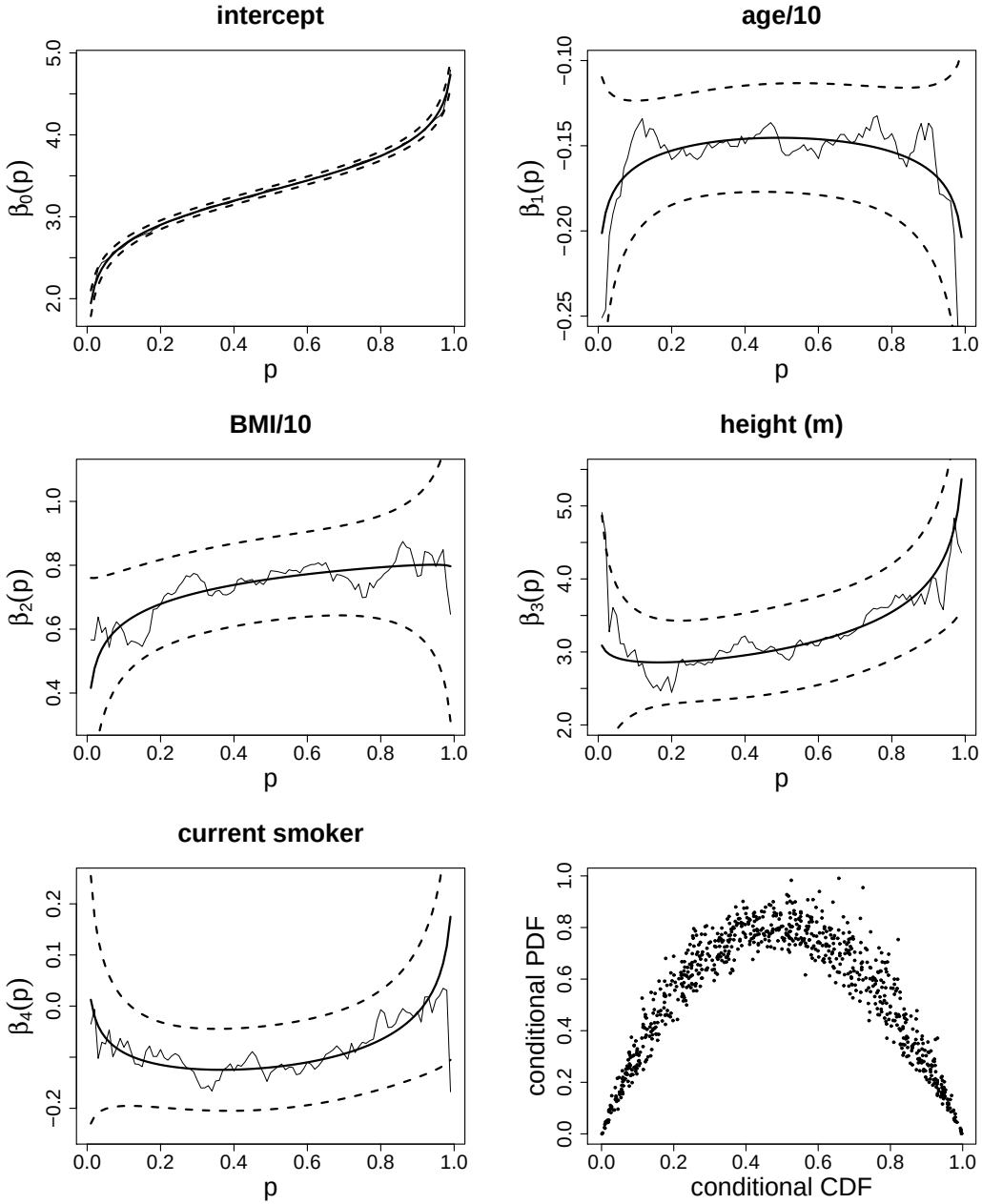
We predict quantiles of inspiratory capacity based on age, BMI, height, and smoking. For different model specifications, we tested

$$H_0 : \text{“the model is correct”}.$$

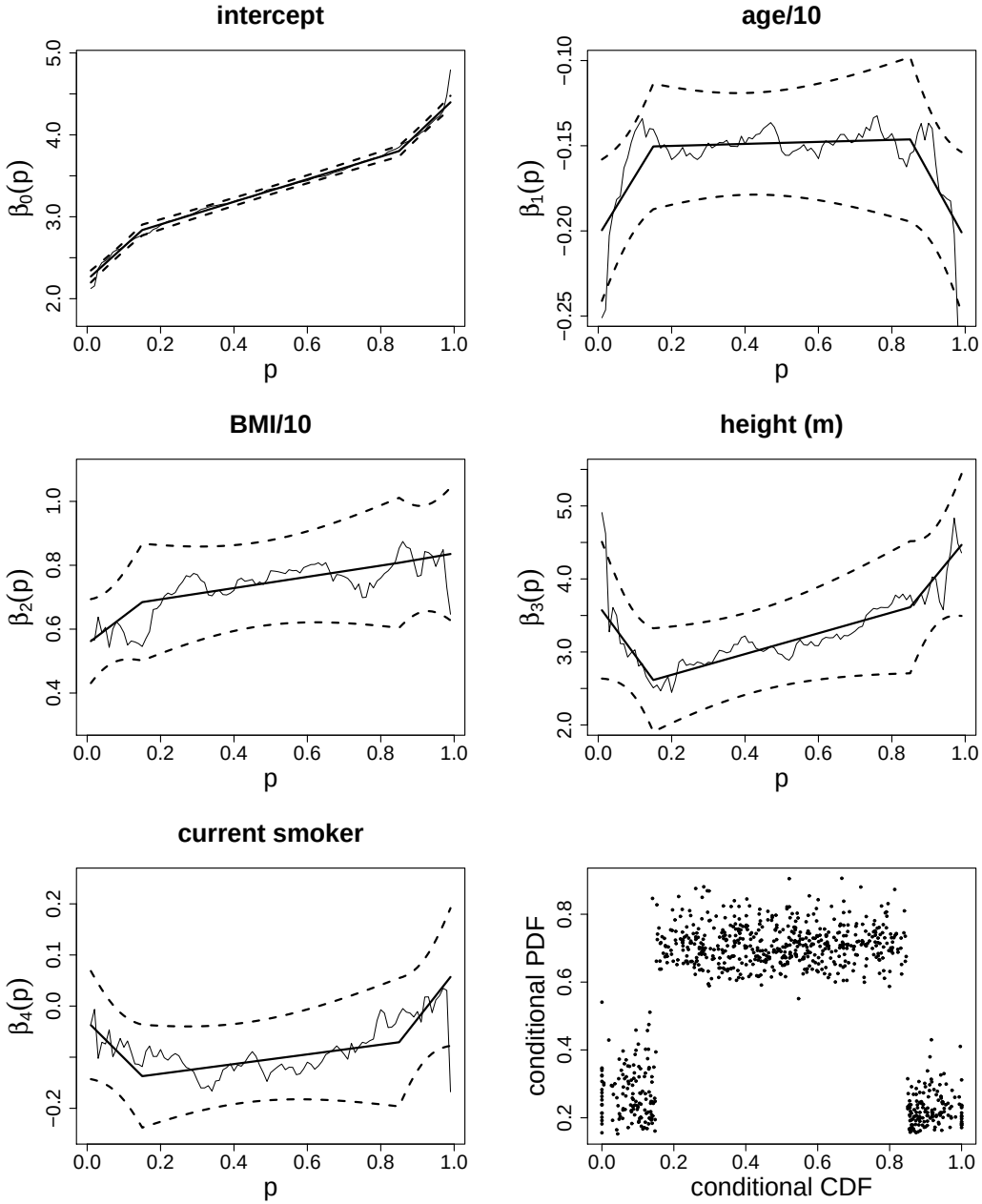
Table: Alternative model specifications

Model	$\boldsymbol{b}(p)$	Loss	P-value H_0
1	p	127.90	0.000
2	p, p^2	127.81	0.000
3	p, p^2, p^3	126.93	0.928
4	$z(p)$ (Normal)	126.98	0.672
5	$\log[p/(1-p)]$	127.01	0.768
6	$\log(p), \log(1-p)$	126.86	0.878
7	piecewise linear	126.98	0.355

Model 6


intercept
β0(p)
p
age/10
β1(p)
p
BMI/10
β2(p)
p
height (m)
β3(p)
p
current smoker
β4(p)
p
conditional PDF
conditional CDF

Model 7


intercept
β0(p)
p
age/10
β1(p)
p
BMI/10
β2(p)
p
height (m)
β3(p)
p
current smoker
β4(p)
p
conditional PDF
conditional CDF

Interpretation

Based on model 6, the coefficient function of smoking is

$$\hat{\beta}_4(p) = -0.21 - 0.05\log(p) - 0.08\log(1-p)$$

You can interpret $\hat{\boldsymbol{\theta}}$ (which may not be simple) or directly obtain $\hat{\boldsymbol{\beta}}(p) = \boldsymbol{\beta}(p \mid \hat{\boldsymbol{\theta}})$. For example,

$$\hat{\beta}_4(0.5) = -0.21 - 0.05\log(0.1) - 0.08\log(1-0.1) = -0.09.$$

Censored and truncated data

- Using the same working principle, we can apply QRCM to censored and truncated data
- Actually, it is rather simple
- In comparison, standard QR with censored and truncated data is a sort of statistical drama

Intuition:

- QR has a nuisance parameter: the entire CDF
- Instead, QRCM has none

Longitudinal data

Assume U_{it} and V_i are independent $U(0,1)$ variables. Define

$$Y_{it} = \boldsymbol{x}_{it}^{\mathrm{T}}\boldsymbol{\beta}(U_{it}) + \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\gamma}(V_i)$$

- $\boldsymbol{x}_{it}$ = level-1 covariates, with associated parameter $\boldsymbol{\beta}(\cdot)$
- $\boldsymbol{z}_i$ = level-2 covariates, with associated parameter $\boldsymbol{\gamma}(\cdot)$
- A two-level quantile function
- $\boldsymbol{\alpha}_i = \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\gamma}(V_i)$ is an individual-level intercept
- Level 1: $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}(u)$ = quantile function of $Y - \boldsymbol{\alpha}$
- Level 2: $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\gamma}(v)$ = quantile function of $\boldsymbol{\alpha}$

Longitudinal data (cont.)

- Relax level-2 normality assumptions
- Level-1 and level-2 parameters are treated equally
- A new form of penalised fixed-effects estimator

Quantile crossing

- A certain parametrisation can be used to avoid crossing
- The parametric structure itself prevents crossing
- Constrained optimisation: $\min \overline{L}(\boldsymbol{\theta})$ s.t. $Q'(p \mid \boldsymbol{x}, \boldsymbol{\theta}) > 0$

Count data

- Parametric model = implicit jittering
- "Smooth away" the points of mass

Variables selection

- You can apply lasso
- Simultaneous selection on “all” quantiles

M-quantiles

- An easy generalisation

A very general objective function

$$L(\boldsymbol{\theta}) = \int_0^1 \sum_{i=1}^{n} w(p)(p - \omega_i(p))(y_i - Q(p \mid \boldsymbol{x}_i, \boldsymbol{\theta}))\mathrm{d}p$$

- Fit any quantile function (e.g., that of a linear or Poisson regression model, or a Cox model)
- Nothing to do with quantile regression!
- Intuition: although parameters might not represent quantiles, they *must* be functions of the quantiles!
- Assign a different weight to each quantile
- A very general robust estimator

Computation: done!

- qrcm: quantile regression coefficients modelling; censored or truncated data and longitudinal responses; quantile crossing.
- Mqrcm: M-quantiles.
- qrcmNP: nonlinear and penalised models.
- Qest: robust estimation and regression with parametric quantile functions.

Examples with qrcm

```
library(qrcm)

iqr(y ~ x) # cross-sectional data
iqr(Surv(time, event) ~ x) # censored
iqr(Surv(start, stop, event) ~ x) # and truncated
iqrL(y ~ x, id) # longitudinal data

summary(model) # model summary
plot(model) # plot coefficient functions
test.fit(model) # goodness-of-fit test
predict(model) # prediction and simulation
```

Computation (cont.)

```
# The 'formula.p' argument

~ p + I(p^2) + I(p^3) # a cubic function
~ I(cos(pi)) + I(sin(pi)) # a trigonometric function
~ I(log(p)) + I(log(1 - p)) # asymmetric logistic
```

- some entries of $\boldsymbol{\theta}$ can be constrained to be zero
- for longitudinal data, you need two 'formula.p' arguments

Conclusions

- An alternative to PDF (likelihood) modelling
- Smooth objective function
- More efficient and parsimonious than QR
- Identification! (problems with latent variables or missing data)

Further applications and generalisations:

- Interval-censored data (work in progress)
- Varying coefficients
- A general theory on extreme estimation?

References

Frumento P, Bottai M (2016). "Parametric modeling of quantile regression coefficient functions". *Biometrics*, 72(1), pp.74-84 .

Frumento P, Bottai M (2017). "Parametric modeling of quantile regression coefficient functions with censored and truncated data". *Biometrics*, 73(4), pp.1179-1188.

Frumento P, Salvati N (2019). "Parametric modelling of M-quantile regression coefficient functions with application to small area estimation". *Journal of the Royal Statistical Society, Series A*.

References (cont.)

Frumento P, Bottai M, and Fernández-Val I (2021). "Parametric modeling of quantile regression coefficient functions with longitudinal data". *Journal of the American Statistical Association, 116(534), pp 783-797*.

Sottile G, Frumento P (2021). "Parametric estimation of non-crossing quantile functions". *Statistical Modelling (online ahead of print), https://doi.org/10.1177/1471082X211036517*.

Sottile G, Frumento P (2022). "Robust estimation and regression with parametric quantile functions". *Computational Statistics and Data Analysis, 171, 107471.* .

R packages

Frumento P (2021). qrcm: Quantile Regression Coefficients Modeling. V.3.0, https://CRAN.R-project.org/package=qrcm

Frumento P (2021). Mqrcm: M-Quantile Regression Coefficients Modeling. V.1.2, https://CRAN.R-project.org/package=Mqrcm

Sottile G (2021). qrcmNP: Nonlinear and Penalized Quantile Regression Coefficients Modeling. V.0.2.0, https://CRAN.R-project.org/package=qrcmNP

Sottile G, Frumento P (2022). Qest: Quantile-Based Estimator. V.1.0.0, https://CRAN.R-project.org/package=Qest