Quantile regression

$\binom{X_i}{Y_i}$ iid $\quad Y_i \approx \beta' X_i + \varepsilon_i$ for ε_i centred, homoscedastic $\quad , \tau \in (0,1)$

$$\hat{\beta}_n(\tau) = \underset{b \in \mathbb{R}^p}{\arg\min} \frac{1}{n} \sum \rho_\tau(Y_i - b'X_i) \quad \text{for } \rho_\tau(x) = \tau x\, I[x>0] + (1-\tau)(-x) I[x<0]$$

Identifies $\beta_X(\tau) = \underset{b \in \mathbb{R}^p}{\arg\min}\, E \rho_\tau(Y - b'X)$

for each X, $E\rho_\tau(Y - F^{-1}_{Y|X}(\tau)) \leq E\rho_\tau(Y - b'X) \quad \forall b$

→ if $F^{-1}_{Y|X}(\tau)$ conditional ~~median~~ quantile of Y given X is linear, we estimate this linear function.

∘ Regression quantiles are equivariant for monotone transforms of Y.

R package quantreg, function rq works analogously as lm for OLS models where conditional expectation is modelled

Inference: $\sqrt{n}(\hat{\beta}_n(\tau) - \beta_X(\tau)) \xrightarrow{d} N_p(0, V)$ $\quad$ [M-estimation and sandwich]

$$V = [E\, XX^T f_{Y|X}(F^{-1}_{Y|X}(\tau))]^{-1} \tau(1-\tau) E XX^T [E\, XX^T f_{Y|X}(F^{-1}_{Y|X}(\tau))]^{-1}$$

If the quantiles given X are indeed linear

$$V = \frac{\tau(1-\tau)}{f^2_\varepsilon(F^{-1}_\varepsilon(\tau))} \frac{1}{n} (E\, XX^T)^{-1}$$

estimated f, for $h_n \to 0$

$$F_\varepsilon \circ F^{-1}_\varepsilon(\tau) = \tau \quad / \partial/\partial\tau$$

$$f_\varepsilon(F^{-1}_\varepsilon(\tau)) \cdot \frac{\partial F^{-1}_\varepsilon(\tau)}{\partial \tau} = 1 \Rightarrow \frac{1}{f_\varepsilon(F^{-1}_\varepsilon(\tau))} = \frac{\partial F^{-1}_\varepsilon(\tau)}{\partial \tau} \approx \frac{\hat{F}^{-1}_{\varepsilon,n}(\tau + h_n) - \hat{F}^{-1}_{\varepsilon,n}(\tau - h_n)}{2h_n}$$

anova.rq: • for one model with different τ ⇒ test of the equality of slopes (quantiles are indeed linear, i.e. the true model is valid, quantiles do not cross each other)

• for nested models with a single τ ⇒ submodel test

+ example 41 from the lecture notes – two sample problem

R script