A Gentle Introduction to Quantile Regression

Stéphane Bonhomme

January 2008

Motivation

We have a sample (iid) of (y_i, x_i), and we want to describe the relation between y_i and x_i.

Classical regression theory focuses on the conditional expectation function $\mathbb{E}(y_i|x_i)$.

$\Rightarrow$ focus on means, on average effects.

In many contexts, describing the full distribution of y_i given x_i is of great interest, e.g. because of the presence of heterogeneity (in preferences, technology, ability...).

This what quantile regression does.

Unconditional quantiles

Denote as $F(y_i)$ the cdf of y_i.

- The median of y_i, that we denote as $\text{med}(y_i)$, solves

$$F\left(\text{med}(y_i)\right) = \frac{1}{2}.$$

$\Rightarrow$ 1/2 observations below, 1/2 above.

- The τth quantile of y_i solves

$$F\left(q_\tau(y_i)\right) = \tau.$$

$\Rightarrow$ τ observations below, $1 - \tau$ above.

The set of $q_\tau(y_i)$, $\tau \in [0, 1]$, describes the full distribution of y_i.

Conditional quantiles

Denote as $F(y_i|x_i)$ the conditional cdf of y_i given x_i.

- Conditional median

$$F\left(\text{med}(y_i|x_i)|x_i\right) = \frac{1}{2}.$$

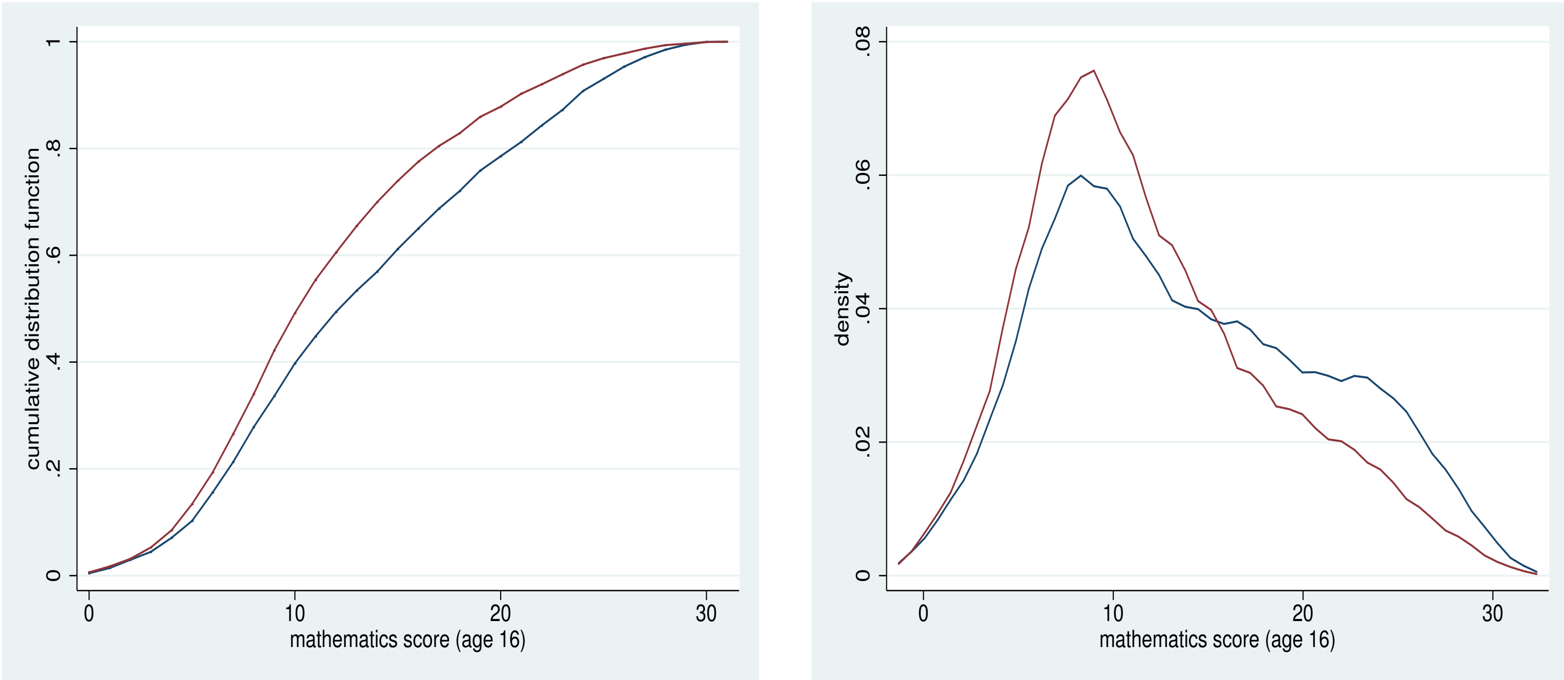
- Conditional τth quantile

$$F\left(q_\tau(y_i|x_i)|x_i\right) = \tau.$$

Quantile regression allows to estimate $q_\tau(y_i|x_i)$, for all $\tau \in [0,1]$.

Example 1: Selective (blue) and non selective (red) schools

Score in mathematics at age 16



Note: NCDS data, England and Wales, 1974.

Example 1: Average and quantile effects

Controls: parental, school, local characteristics, + the first principal component of lagged test scores.

- *reg math16 selective controls*

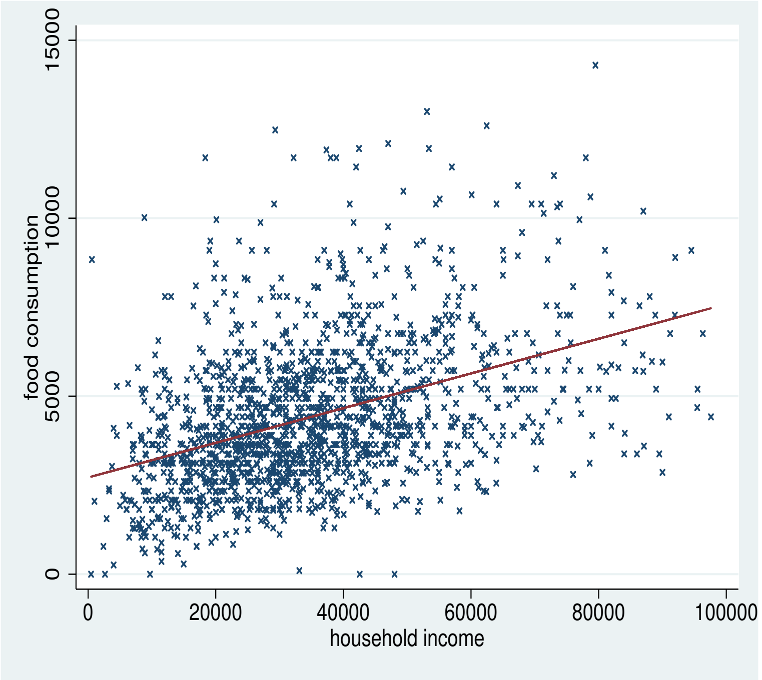
Average effect: .56 (.13).

- *qreg math16 selective controls, q(*10*)*

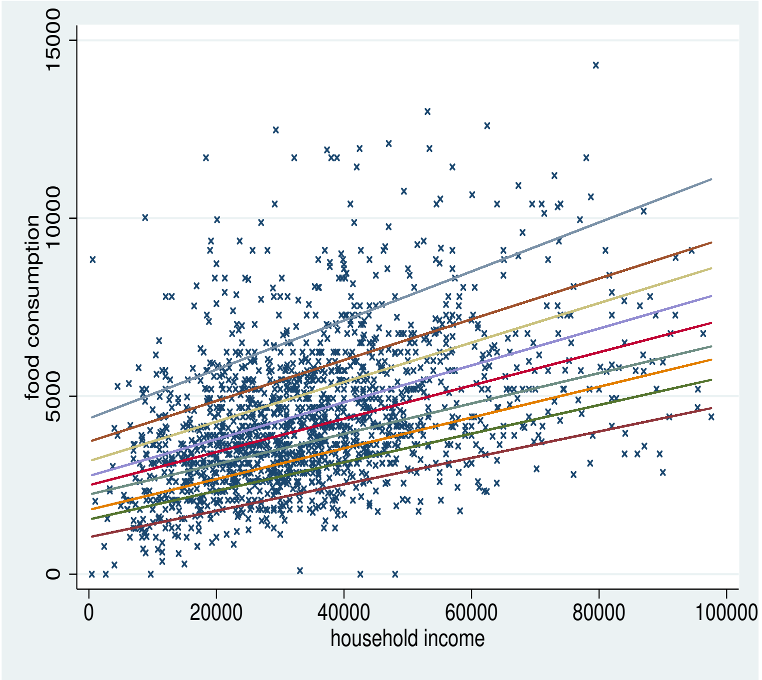
Effects at the different deciles:

Decile	D1	D2	D3	D4	D5	D6	D7	D8	D9
Effect	.29 (.21)	.21 (.20)	.39 (.20)	.54 (.16)	.67 (.16)	.52 (.15)	.64 (.21)	.69 (.23)	.51 (.21)

Example 2: Food consumption and household income



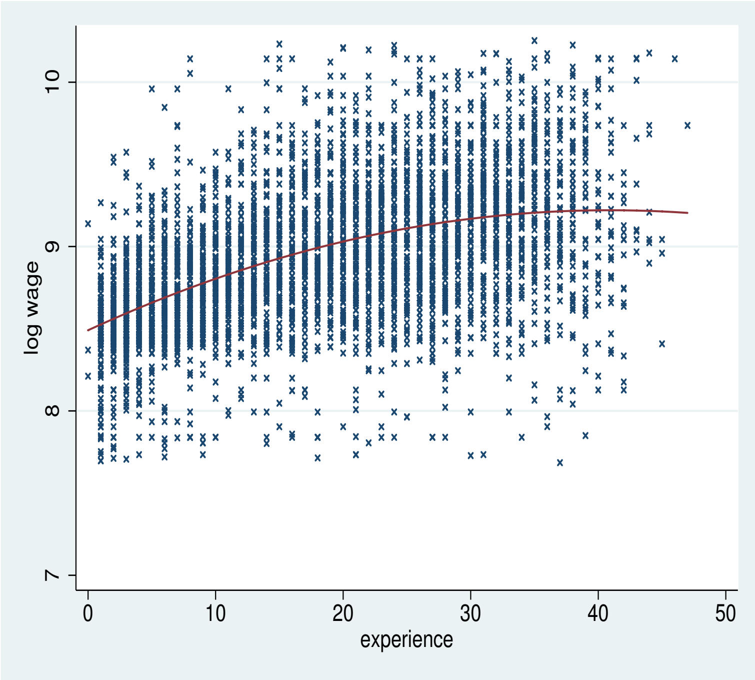
regression line



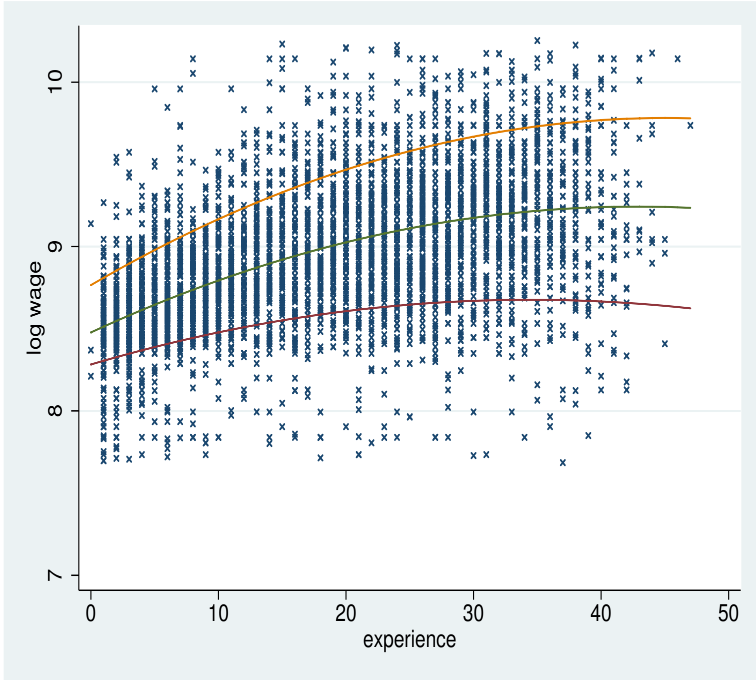
quantile regression lines (deciles D1-D9)

Note: PSID data for 1985, male workers with college education.

Example 3: Log-wages and experience (quadratic)



regression curve



quantile regression curves
(deciles D1, D5 and D9)

Note: French labor force survey for 2000, male workers without a college degree.

Outline

- Motivation
- Quantiles in regression analysis
- Characterizing quantiles
- The quantile regression estimator
- Important issues
- Conclusion

Quantiles in regression analysis

The location-scale model

Consider the following additive model with conditional heteroskedasticity:

$$y_i = \mu(x_i; \beta) + \sigma(x_i; \gamma) u_i$$

where $u_i | x_i \sim G$, independent of x_i.

In this model:

$$q_\tau (y_i | x_i) = \mu(x_i; \beta) + \sigma(x_i; \gamma) G^{-1}(\tau).$$

Proof

The conditional cdf of y_i given x_i is

$$F(y_i|x_i) = G\left(\frac{y_i - \mu(x_i;\beta)}{\sigma(x_i;\gamma)}\right)$$

so:

$$\begin{aligned} F\left(\mu(x_i;\beta) + \sigma(x_i;\gamma)G^{-1}(\tau)|x_i\right) &= G\left(\frac{\mu(x_i;\beta) + \sigma(x_i;\gamma)G^{-1}(\tau) - \mu(x_i;\beta)}{\sigma(x_i;\gamma)}\right) \\ &= G\left(G^{-1}(\tau)\right) = \tau. \end{aligned}$$

This implies

$$q_\tau\left(y_i|x_i\right) = \mu(x_i;\beta) + \sigma(x_i;\gamma)G^{-1}(\tau).$$

Location-scale and quantiles

$$q_\tau\left(y_i|x_i\right) = \mu(x_i;\beta) + \sigma(x_i;\gamma)G^{-1}(\tau).$$

$$\Rightarrow \frac{\partial q_\tau\left(y_i|x_i\right)}{\partial x_i} = \frac{\partial \mu(x_i;\beta)}{\partial x_i} + \frac{\partial \sigma(x_i;\gamma)}{\partial x_i}G^{-1}(\tau).$$

• Conditional homoskedasticity $\Rightarrow$ quantile curves are parallel. So quantile effects and average effects are the same.

• Conditional heteroskedasticity $\Rightarrow$ quantile curves are not parallel. Yet...

Location-scale and quantiles (continued)

In the location-scale model, all parameters share the same monotone behavior in τ.

For instance, imposing linearity of μ and σ we obtain

$$\frac{\partial q_{\tau}\left(y_i|x_i\right)}{\partial x_i}=\beta+G^{-1}(\tau)\gamma.$$

$\Rightarrow$ restrictive.

Alternative:

$$q_{\tau}(y_i|x_i)=x_i'\beta_{\tau}.$$

Linear in x_i, but different effects of τ on different outcomes. Example: the union wage effect in Chamberlain (1994), the determinants of birth weight in Abrevaya (2001).

Characterizing quantiles

Quadratic loss

The conditional expectation maximizes the quadratic loss

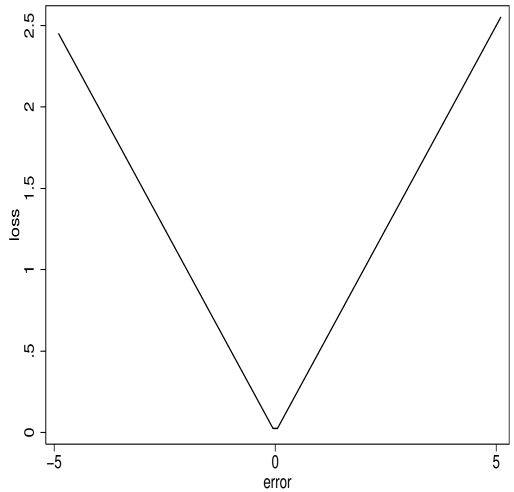
$$\mathbb{E}\left(y_i|x_i\right) = \underset{h(x_i)}{\text{argmin}} \quad \mathbb{E}\left(\left(y_i - h(x_i)\right)^2\right).$$

This property is the basis for the Ordinary and Nonlinear Least Squares (OLS and NLS) estimators.

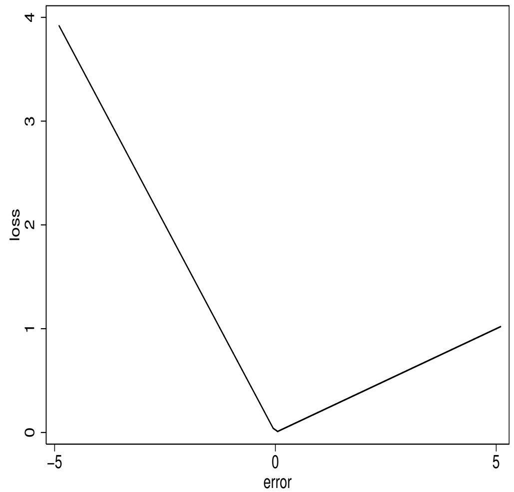
Quantile regression can be similarly motivated, provided that we consider other, possibly asymmetric, loss functions.

The functions ρ_τ, $\tau \in [0,1]$

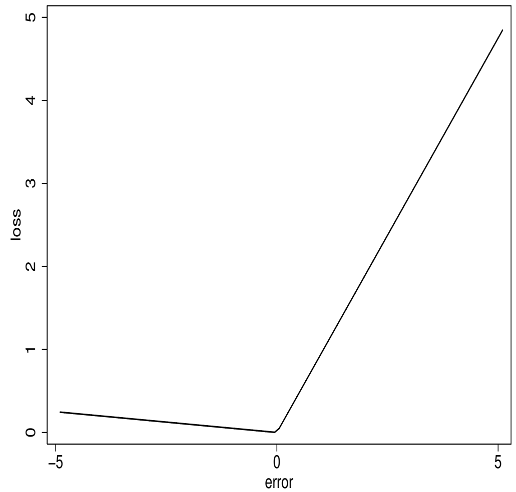
$$\rho_\tau(u) = \begin{cases} \tau u & \text{if} \quad u \geq 0 \\ -(1-\tau)u & \text{if} \quad u \leq 0 \end{cases}$$



$\tau = .50$



$\tau = .20$



$\tau = .95$

Fundamental result

$$q_\tau(y_i|x_i) = \underset{h(x_i)}{\text{argmin}} \quad \mathbb{E}\left(\rho_\tau\left(y_i - h(x_i)\right)\right).$$

In particular:

$$\text{med}\,(y_i|x_i) = q_{\frac{1}{2}}(y_i|x_i) = \underset{h(x_i)}{\text{argmin}} \quad \mathbb{E}\left(|y_i - h(x_i)|\right).$$

$\Rightarrow$ basis for consistent estimation of quantile functions.

Proof (unconditional case)

$$\mathbb{E}\left(\rho_\tau\left(y_i - h\right)\right) \;=\; \tau \int_h^{+\infty} (y-h) f(y) dy + (\tau - 1) \int_{-\infty}^{h} (y-h) f(y) dy.$$

Evaluating the FOC at $h = q$:

$$-\tau \int_q^{+\infty} f(y) dy - (\tau - 1) \int_{-\infty}^{q} f(y) dy \;=\; 0.$$

That is:

$$-\tau(1 - F(q)) - (\tau - 1) F(q) \;=\; 0.$$

Hence:

$$F(q) = \tau, \quad \Rightarrow \quad q = q_\tau.$$

Important: F is differentiable (around q_τ), i.e. y_i is continuous.

Linear quantile functions

Assume that

$$q_\tau\left(y_i|x_i\right) = x_i'\beta_\tau.$$

Then:

$$\beta_\tau = \underset{b}{\text{argmin}} \quad \mathbb{E}\left(\rho_\tau\left(y_i - x_i'b\right)\right).$$

In particular

$$\beta_{\frac{1}{2}} = \underset{b}{\text{argmin}} \quad \mathbb{E}\left(\left|y_i - x_i'b\right|\right).$$

Remark: if the quantile functions are not linear, $x_i'\beta_\tau$ can be thought of as a linear projection of $q_\tau\left(y_i|x_i\right)$ on x_i in some sense (Angrist, Chernozhukov and Fernandez-Val, 2006).

A random-coefficient interpretation

The linear quantile regression model is

$$q_\tau\left(y_i|x_i\right) = x_i'\beta_\tau,$$

for all $\tau \in [0,1]$. So, to simulate y_i we can proceed as follows:

1. Draw $u_i \sim \mathcal{U}[0,1]$.

2. Compute $y_i = x_i'\beta_{u_i}$.

$\Rightarrow$ The components of β_τ can be thought of as random coefficients, with a one-factor structure.

A method-of-moments interpretation

$$\tau = F\left(q_\tau\left(y_i|x_i\right)|x_i\right) = \mathbb{E}\left(\mathbf{1}\{y_i \leq q_\tau\left(y_i|x_i\right)\}|x_i\right).$$

$$\Rightarrow \quad \mathbb{E}\left(\mathbf{1}\{y_i \leq q_\tau\left(y_i|x_i\right)\} - \tau|x_i\right) = 0.$$

$$\Rightarrow \quad \mathbb{E}\left(\left[\mathbf{1}\{y_i \leq q_\tau\left(y_i|x_i\right)\} - \tau\right] g(x_i)\right) = 0 \text{ for all functions } g.$$

In particular, in the linear quantile regression model

$$\mathbb{E}\left(\left[\mathbf{1}\{y_i \leq x_i'\beta_\tau)\} - \tau\right] g(x_i)\right) = 0.$$

This interpretation is important in the case of quantile IV.

The quantile regression estimator

Definition

The analogy principle suggests considering:

$$\widehat{\beta}_\tau = \underset{b}{\text{argmin}} \quad \sum_{i=1}^{N} \rho_\tau \left(y_i - x_i' b \right).$$

This is the quantile regression estimator of Koenker and Basset (1978).

In particular, for the median

$$\widehat{\beta}_{\frac{1}{2}} = \underset{b}{\text{argmin}} \quad \sum_{i=1}^{N} \left| y_i - x_i' b \right|.$$

This is the Least Absolute Deviations (LAD) estimator.

Computation

No explicit solution (unlike OLS) and non-differentiable objective function (unlike regular NLS). Standard optimization methods (e.g., Newton) cannot be used.

The key observation is that the minimization problem can be seen as a linear program.

$=$ optimization of a linear function subject to linear constraints.

The minimization problem as a linear program

$$\min_{b \in \mathbb{R}^K} \quad \sum_{i=1}^{N} \rho_\tau \left(y_i - x_i' b \right).$$

is equivalent to

$$\begin{aligned} \min \quad & \sum_{i=1}^{N} (\tau u_i + (1-\tau) v_i) \\ s.t. \quad & y_i - x_i' b = u_i - v_i, \quad u_i \geq 0, v_i \geq 0, \quad b \in \mathbb{R}^K. \end{aligned}$$

One can restrict the attention to the b's that are interpolating K observations (y_i, x_i).

The minimization in practice

Several computationally efficient optimization methods have been developed, mostly based on the simplex algorithm.

$\Rightarrow$ very fast

Example: 3 regressors (expe, expe^2, constant). STATA *qreg* command.

Sample size N	100	1000	10000	100000
Time (seconds)	.02	.03	.30	60.53

Properties

The intuition for consistency of the quantile regression estimator comes from the analogy principle. Technical conditions are given in Chapter 4 of Koenker (2005).

In addition, the estimator is root-N consistent and asymptotically normal under stronger conditions.

The difficulty for proving the asymptotic normality comes from the nondifferentiability of the (sample) objective function.

Asymptotic normality of sample quantiles

(Heuristic) The first-order condition for $\widehat{q}_\tau$ is

$$\widehat{F}(\widehat{q}_\tau) - \tau = 0,$$

where $\widehat{F}(y) = N^{-1}\sum_{i=1}^N \mathbf{1}\{y_i \leq y\}$ is the empirical cdf of y_i.

So:

$$0 = \widehat{F}(\widehat{q}_\tau) - \tau = \widehat{F}(q_\tau) - \tau + \widehat{F}(\widehat{q}_\tau) - \widehat{F}(q_\tau).$$

The difference with usual asymptotic arguments is that $\widehat{F}$ is not differentiable (not even continuous).

The key observation is that the limit of $\widehat{F}$ when N tends to infinity, which is F by the Glivenko-Cantelli theorem, is differentiable.

Asymptotic normality of sample quantiles (continued)

So:

$$\begin{aligned} \tau - \widehat{F}(q_\tau) &\approx F(\widehat{q}_\tau) - F(q_\tau) \\ &\approx f(q_\tau)(\widehat{q}_\tau - q_\tau) \end{aligned}$$

It follows that

$$\sqrt{N}(\widehat{q}_\tau - q_\tau) \approx \sqrt{N}\left(\frac{\tau - \widehat{F}(q_\tau)}{f(q_\tau)}\right).$$

But $\mathbb{E}\left(\widehat{F}(q_\tau)\right) = \tau$, and Var $\left(\widehat{F}(q_\tau)\right) = \tau(1-\tau)$. So:

$$\sqrt{N}(\widehat{q}_\tau - q_\tau) \underset{d}{\rightarrow} \mathcal{N}\left(0, \frac{\tau(1-\tau)}{f(q_\tau)^2}\right).$$

$\Rightarrow$ to estimate a quantile precisely, we need enough density around it!

“Sandwich” formula for the quantile regression estimator

In the case of the quantile regression estimator, the population objective function is differentiable with respect to the parameter b.

$\Rightarrow$ similar strategy and result.

$$\sqrt{N}\left(\widehat{\beta}_\tau - \beta_\tau\right) \underset{d}{\rightarrow} \mathcal{N}\left(0, \tau(1-\tau) D_1^{-1} D_0 D_1^{-1}\right).$$

$$D_0 = \lim_{N\rightarrow\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(x_i x_i'\right); \quad D_1 = \lim_{N\rightarrow\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(f_{y_i|x_i}(x_i'\beta_\tau) x_i x_i'\right).$$

Standard errors estimation

Two strategies:

- Estimation based on the theoretical "sandwich" formula.

Hendricks-Koenker (1991), and Powell (1991).

- (Nonparametric) Bootstrap. Practical if qreg is fast enough.

$\Rightarrow$ Tests. Example: test that all quantile lines are parallel.

Remark: STATA's standard errors in *qreg* are often underestimates of the true one. *bsqreg* is an alternative.

Properties: robustness

The median is thought to be more robust to outliers than the mean.

A similar intuition applies to median– or quantile– regression vis à vis OLS.

Example: wage on experience; 5935 observations. I increase the logwage of the first observation to 1000 (mean log wage *ex-ante* ≈ 9).

$\Rightarrow$ OLS decreases from .019 to $-.005$; median regression unchanged (.021).

Important issues

Equivariance

Let h be a nondecreasing function. Then

$$q_\tau\left(h(y_i)|x_i\right) = h\left(q_\tau\left(y_i|x_i\right)\right).$$

$\Rightarrow$ Quantiles are equivariant to monotone transformations.

Obvious, as

$$P(y_i \leq q) = P(h(y_i) \leq h(q)).$$

Very important! Contrasts with (if h is nonlinear):

$$\mathbb{E}\left(h(y_i)|x_i\right) \neq h\left(\mathbb{E}\left(y_i|x_i\right)\right).$$

Application 1: transformation models

Example (Box-Cox transformation):

$$q_{\tau}\left(\frac{y_i^{\lambda}-1}{\lambda}|x_i\right)=x_i'\beta_{\tau}.$$

Quantile regression of $\frac{y_i^{\lambda}-1}{\lambda}$ on $x_i \Rightarrow \widehat{\beta}_{\tau}$.

Then, one is justified to interpret

$$\left(1+\lambda x_i'\widehat{\beta}_{\tau}\right)^{\frac{1}{\lambda}}$$

as an estimate of the τ's conditional quantile of y_i given x_i.

Application 2: censoring

Assume that

$$q_{\tau}\left(y_i^*|x_i\right) = x_i'\beta_{\tau},$$

but that we observe only $y_i = \max(y_i^*, 0)$.

Then equivariance implies, using an idea of Powell (1986)

$$q_{\tau}\left(y_i|x_i\right) = \max\left(x_i'\beta_{\tau}, 0\right).$$

So it is natural to estimate β_{τ} as

$$\widehat{\beta}_{\tau} = \underset{b}{\operatorname{argmin}} \quad \sum_{i=1}^{N} \rho_{\tau}\left(y_i - \max(x_i'b, 0)\right).$$

Oops! Quantile lines cross

For given x_i, $q_\tau(y_i|x_i)$ should be nondecreasing in τ. However, the quantile regression estimates $x_i'\widehat{\beta}_\tau$ can fail to be, because:

- they are estimates (N is finite)

- the linear quantiles specification is incorrect.

$\Rightarrow$ The implied conditional cdf makes no sense. For instance it is not clear how to simulate the model based on the estimated parameters.

Koenker (2005) argues that quantile crossing is most often confined to "outlying regions of the design space" (p. 56)...

Finally... a bad idea

I consider wages and experience. To estimate the 95% quantile slope, I first keep only the observations that are above the 90%-quantile of log-wages, then I regress log-wages on experience by OLS.

I find a coefficient of .0032 (std error= .0009).

In contrast, the *qreg, q(95)* command yields .027 (std error= .0009).

$\Rightarrow$ Truncating the dependent variable does not work, because we are considering a selected sample!

Conclusion

Quantile regression offers a powerful and practical complement to OLS.

It can allow to estimate effects at different points of the distribution, not only mean effects.

$\Rightarrow$ Do not be afraid to use it!

Question: how to deal with endogeneity? Is there a quantile equivalent to 2SLS?

Answer next week!