Chapter 8. Quantile Regression and Quantile Treatment Effects

Joan Llull

Quantitative Statistical Methods II
Barcelona School of Economics

Introduction

Motivation

So far in this course: **conditional averages** $\mathbb{E}[Y_i|X_i]$.

We may be interested in other characteristics of the **distribution** (e.g. inequality).

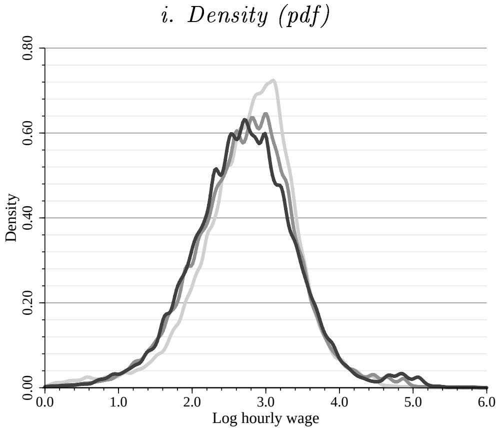
The τth **quantile** of the distribution of Y_i is the value q_τ for which a fraction τ of the population has $Y_i \leq q_\tau$.

$\Rightarrow$ Quantiles **fully characterize** the distribution of Y_i.

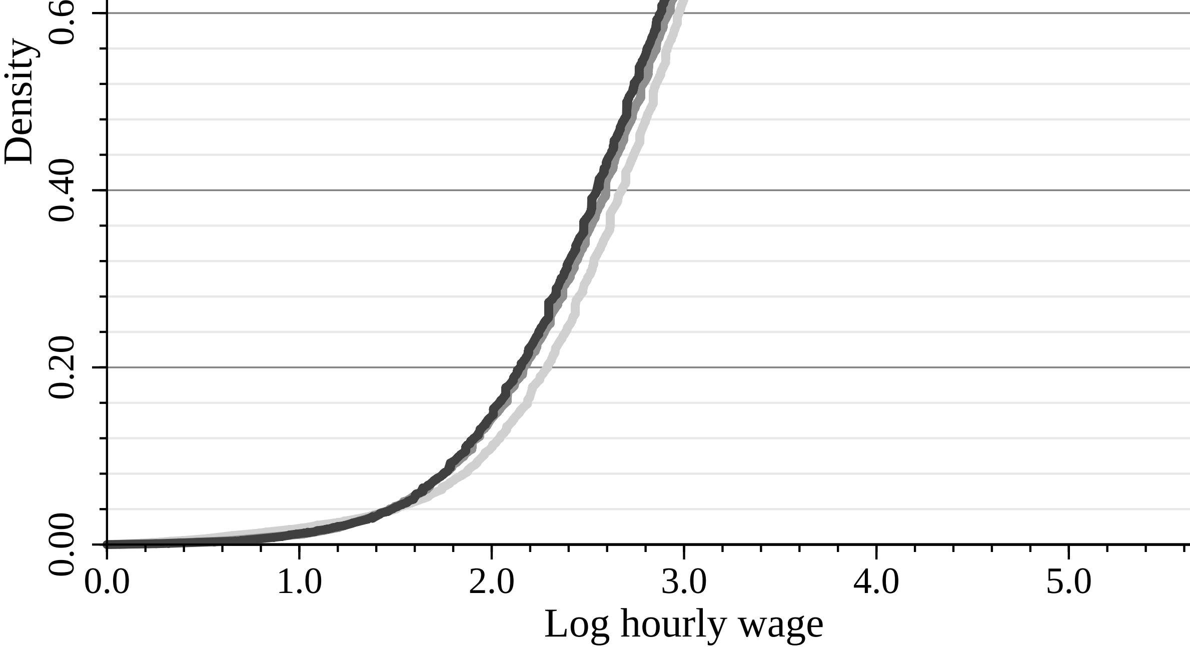
Most popular quantile: the **median** (other popular quantiles?)

Example: Why wage at the 20th percentile is 10-20% larger in 1980 than in subsequent years?

Figure I. – Distribution of U.S. Male Wages (1980-2000)



Note: Light gray: 1980; gray: 1990; dark gray: 2000. Sample restricted to working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. *Data source*: U.S. Census.


Density
0.00
0.20
0.40
0.6
0.0
1.0
2.0
3.0
4.0
5.0
Log hourly wage

Unconditional quantiles

We will introduce the general notation with **unconditional quantiles**.

Let $F(Y_i)$ be the cdf of Y_i. The τ**th quantile** of Y_i, $q_\tau(Y_i)$ solves:

$$F(q_\tau(Y_i)) = \tau \qquad \Leftrightarrow \qquad q_\tau(Y_i) = F^{-1}(\tau),$$

or: the value of Y_i that leaves a fraction τ of observations below and $1-\tau$ above.
(read them in the graph)

The distribution of Y_i is **fully described** by $\{q_\tau(Y_i), \tau \in (0,1)\}$.

Median as a special case: $q_{0.5}(Y_i)$ is the value of Y_i that leaves half of the observations above and half below.

Table I. – Unconditional Quantiles for Wages (1980-2000)

	Percentile:				
Year	10th	25th	50th	75th	90th
1980	1.96	2.41	2.84	3.18	3.50
1990	1.86	2.30	2.76	3.15	3.51
2000	1.83	2.27	2.70	3.15	3.55

Note: Sample restricted to working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. *Data source*: U.S. Census.

Sample quantiles

Consider a random sample $\{Y_1, ..., Y_N\}$. **Two ways** of computing quantiles:

$$\hat{F}_N(r) = \frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}} \mathbb{1}\{Y_i \leq r\} \quad \Leftrightarrow \quad \hat{q}_\tau(Y_i) = \hat{F}_N^{-1}(\tau) \equiv \inf\{r : \hat{F}_N(r) \geq \tau\}.$$

(very costly computationally —why?)

For the second, we introduce the **"check function"**:

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

Then quantiles can be calculated as:

$$\hat{q}_\tau(Y_i) = \arg\min_r \sum_{\mathrm{i=1}}^{\mathrm{N}} \rho_\tau(Y_i - r) = \arg\min_r \sum_{Y_i \geq r} \tau|Y_i - r| + \sum_{Y_i \leq r} (1-\tau)|Y_i - r|.$$

(intuition + no analytical solution/easy computationally + the median + population analogue)

Standard errors

Non-differentiability makes asymptotic results non-trivial.

Asymptotic normality can still be established under suitable conditions.

When this is possible, the resulting asymptotic distribution is:

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

where $f(\cdot)$ is the pdf of the distribution $F(\cdot)$. (where are quantiles more precise?)

In practice: standard errors are **bootstrapped** whenever it is feasible.

Nonparametric conditional quantiles

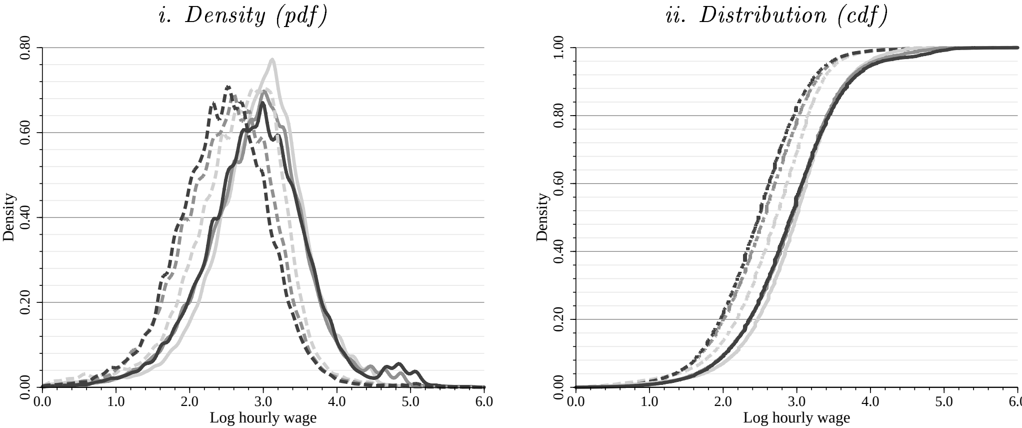
The quantile regression model below is **semiparametric**.

Before: **nonparametric conditional** quantiles as a motivation.

Is it the increasing wage inequality the result of an increase in **education**?

Some insights looking at $q_\tau(Y_i|X_i)$ instead of $q_\tau(Y_i)$!

Figure II. – U.S. Wages Distribution (college vs non-college)



Note: Solid: college; dashed: noncollege. Light gray: 1980; gray: 1990; dark gray: 2000. Sample restricted to working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. Individuals with high school or less are considered as noncollege, whereas individuals with some college or a college degree are considered as college educated. *Data source*: U.S. Census.

(a note on terminology + why not always nonparametric?)

Quantile Regression

Conditional quantiles (again)

Same notation: replace marginal cdf $F(Y_i)$ for its **conditional** counterpart $F(Y_i|X_i)$.

The τ**th conditional quantile** of Y_i, $q_\tau(Y_i|X_i)$ solves:

$$q_\tau(Y_i|X_i) = F^{-1}(\tau|X_i).$$

The **population quantile** problem is:

$$q_\tau(Y_i|X_i) = \arg\min_{q(X_i)} \mathbb{E}[\rho_\tau(Y_i - q(X_i))].$$

- ***Nonparametric case***: $q_\tau(Y_i|X_i)$ unrestricted.
- ***Quantile regression***: some linearity assumptions.

The quantile regression model

First introduced by **Koenker and Basset (1978)**

We will see two different models:

- ***Location-scale model***: may not impose linearity, but restrictive in the heterogeneous effects across quantiles.

- ***General quantile regression model***: linearity assumptions, but flexible across quantiles.

The location-scale model

Consider the following 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:

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

In this model, all dependence of Y_i on X_i occurs through **mean translations** — location— ($\mu(X_i;\beta)$) and **variance re-scaling** ($\sigma(X_i;\gamma)$):

$$\frac{\partial q_\tau(Y_i|X_i)}{\partial X_i} = \frac{\partial \mu(X_i;\beta)}{\partial X_i} + \frac{\partial \sigma(X_i;\gamma)}{\partial X_i}G^{-1}(\tau).$$

(why is this restrictive?)

The general quantile regression model

A more **general** model:

$$q_\tau(Y_i|X_i) = X_i'\beta_\tau.$$

It imposes **linearity** but allows for **different effects** on different quantiles.

It can be seen as a **random coefficients** model (everyone has its own β_τ):

$$\beta_\tau = \beta(U_i).$$

Estimation

The estimation is **analogous to the unconditional** case:

$$\widehat{\beta}_\tau = \arg\min_b \sum_{\mathrm{i=1}}^{\mathrm{N}} \rho_\tau(Y_i - X_i'b) = \arg\min_b \sum_{Y_i \geq X_i'b} \tau|Y_i - X_i'b| + \sum_{Y_i \leq X_i'b} (1-\tau)|Y_i - X_i'b|.$$

Particular case of the **median**:

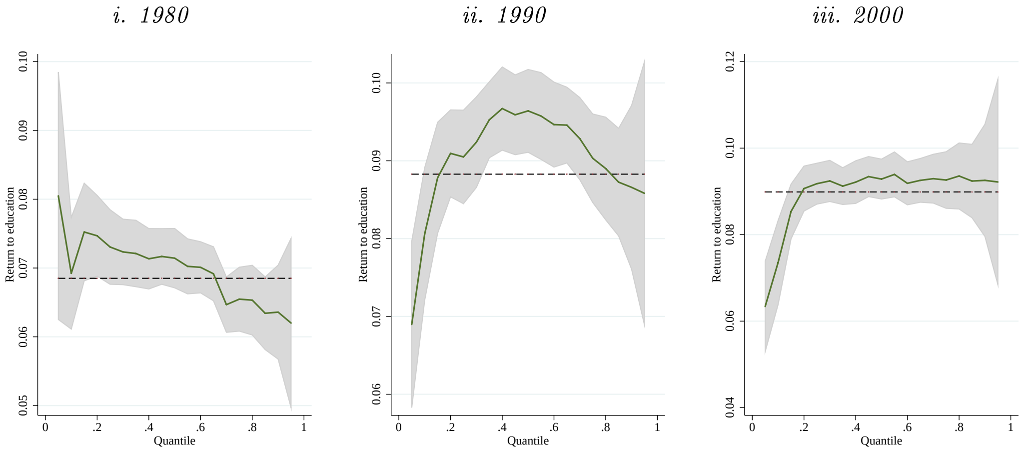
$$\widehat{\beta}_{LAD} \equiv \widehat{\beta}_{0.5} = \arg\min_b \sum_{\mathrm{i=1}}^{\mathrm{N}} |Y_i - X_i'b|.$$

(How does the LAD connect to OLS?)

Example:

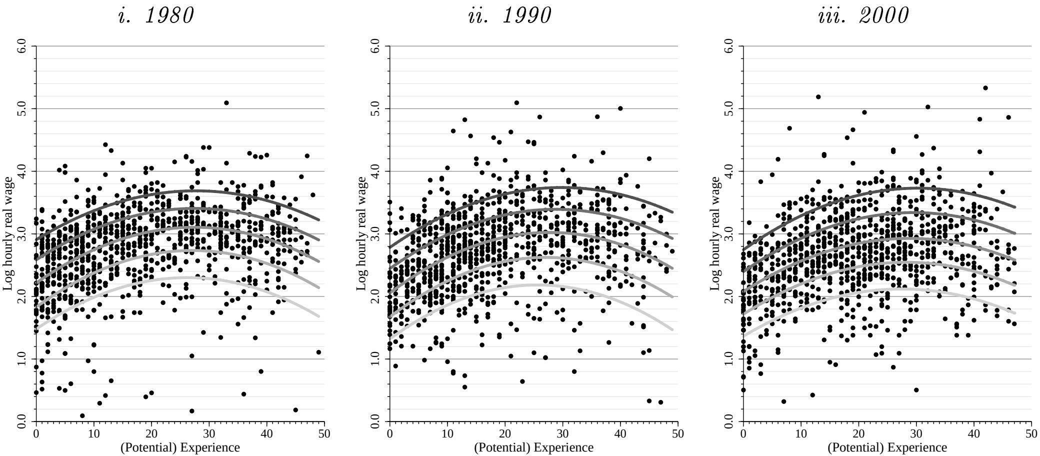
$$q_\tau(\ln W_i|E_i, X_i, X_i^2) = \beta_\tau^{(0)} + \beta_\tau^{(E)} E_i + \beta_\tau^{(X)} X_i + \beta_\tau^{(X^2)} X_i^2.$$

Figure III. – Quantile Regression Coefficients (Education)



Note: Random sample of 10,000/year working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. *Data source*: U.S. Census.

Figure IV. – Quantiles of Wages Conditional on Experience



Note: Quantiles computed with a random sample of 10,000/year working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. The scatter plot depicts a sample of 1,000 observations. *Data source*: U.S. Census.

Quantile regression with censoring

Often we have **censored data** (e.g. **top-coded wages**: we observe $Y_i^* = \min(Y_i, c)$ instead of Y_i).

Averages are affected by censoring, **quantiles** (below the censoring point) are not!

$$q_\tau(Y_i^*|X_i) = \min(X_i'\beta_\tau, c).$$

Hence, using an idea by Powell (1986), we can estimate β_τ as:

$$\widehat{\beta}_\tau^c = \arg\min_b \sum_{\mathrm{i=1}}^{\mathrm{N}} \rho_\tau(Y_i - \min(X_i'b, c)).$$

Quantile Treatment Effects (QTE) Estimator

Quantile treatment effects (QTE)

Quantile treatment effects: introduced by Abadie, Angrist, and Imbens (2002).

The model we are after is: (e.g. college subsidy)

$$q_\tau(Y_i|X_i, D_i, D_{1i} > D_{0i}) = \alpha_\tau D_i + X_i'\beta_\tau.$$

We need an instrument Z_i (that we will assume binary) that picks the **correct group of compliers**. (e.g. distance to college)

Our parameter of interest is:

$$\alpha_\tau = q_\tau(Y_{1i}|X_i, D_{1i} > D_{0i}) - q_\tau(Y_{0i}|X_i, D_{i1} > D_{0i}).$$

Note that:

- $\alpha_\tau \neq q_\tau(Y_{1i}) - q_\tau(Y_{0i})$.
- $\alpha_\tau \neq q_\tau(Y_{1i} - Y_{0i}|X_i, D_{1i} > D_{0i})$.

QTE estimator

We want this:

$$(\alpha_\tau, \beta'_\tau) = \arg\min_{(a,b')} \mathbb{E}[\rho_\tau(Y_i - aD_i - X'_i b) | D_{1i} > D_{0i}].$$

We can use Abadie (2003) result:

$$\mathbb{E}[g(Y_i, X_i, D_i) | D_{1i} > D_{0i}] = \frac{\mathbb{E}[\kappa_i g(Y_i, X_i, D_i)]}{\mathbb{E}[\kappa_i]},$$

where:

$$\kappa_i \equiv 1 - \frac{D_i(1 - Z_i)}{1 - \Pr(Z_i = 1 | X_i)} - \frac{(1 - D_i) Z_i}{\Pr(Z_i = 1 | X_i)}.$$

Hence:

$$(\alpha_\tau, \beta'_\tau) = \arg\min_{(a,b')} \mathbb{E}[\kappa \rho_\tau(Y_i - aD_i - X'_i b)].$$

QTE estimator in practice

κ_i is negative for never-takers and always-takers. **Law of iterated expectations**:

$$(\alpha_\tau, \beta'_\tau) = \arg\min_{(a,b')} \mathbb{E}[\mathbb{E}[\kappa_i|Y_i, X_i, D_i]\rho_\tau(Y_i - aD_i - X_i'b)],$$

where:

$$\mathbb{E}[\kappa_i|Y_i, X_i, D_i] = 1 - \frac{D_i(1 - \mathbb{E}[Z_i|Y_i, X_i, D_i = 1])}{1 - \Pr(Z_i = 1|X_i)} - \frac{(1 - D_i)\,\mathbb{E}[Z_i|Y_i, X_i, D_i = 0]}{\Pr(Z_i = 1|X_i)}.$$

Two-step procedure:

1. **Probit** on $D_i = 0$ and $D_i = 1$ subsamples $\Rightarrow \hat{\mathbb{E}}[Z_i|Y_i, X_i, D_i]$; **Probit** with whole sample $\Rightarrow \Pr(Z_i = 1|X_i)$. **Construct** $\hat{\mathbb{E}}[\kappa_i|Y_i, X_i, D_i]$.
2. Estimate the quantile regression model with the standard procedure (e.g. with `qreg`) using these predicted kappas as **weights**.