Quantile Regression and R-Program

Cliff J. Huang
Emeritus Professor of Economics
Vanderbilt University

Yung-Lieh Yang
Professor of Finance
Ling Tung University

Outline

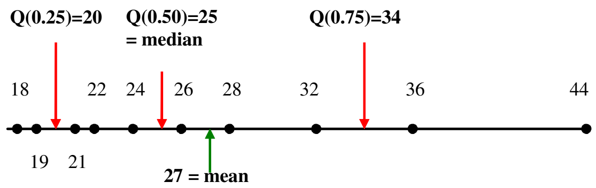
- Mean, Quantiles, and Distribution
- Quantile Regression
- Fundamentals of Quantile Estimation
- Computation of Quantile Regression: R-Program

Mean, Quantiles, and Distribution

1. Distribution of a Random Variable

- Example: Starting Salary of 10 Employees (in thousands of NT$)

  Y = 18, 19, 21, 22, 24, 26, 28, 32, 36, 44



❖ Mean = 27

❖ Quartiles:

First Quartile (25 Percentile) = 20

Second Quartile (50 Percentile) = 25 = Median

Third Quartile (75 Percentile) = 34

❖ Distribution Skewness

Symmetric (No Skewness):

Median = Mean

Median – First Quartile = Third Quartile – Median

Positive Skewness (Skew to the Right)

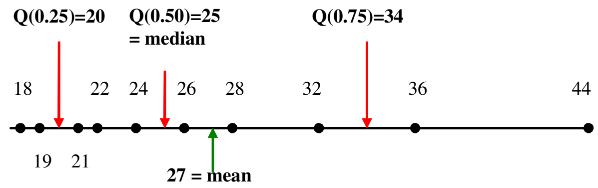
Median < Mean

Median – First Quartile < Third Quartile – Median

Negative Skewness (Skew to the Left)

Median > Mean

Median – First Quartile > Third Quartile – Median



- Quantiles of order τ: $Q(\tau), \quad 0 \le \tau \le 1$

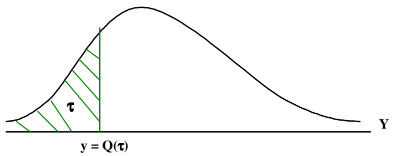
❖ $Q(0.20) = 19; \quad Q(0.25) = 20;$
$Q(0.50) = 25; \quad Q(0.75) = 34;$
$Q(0.90) = 36; \quad Q(1.00) = ?$

❖ Quantile as a Ranking:
$Q(\tau) < Q(\lambda) \quad \text{if } \tau < \lambda$

2. Quantile and Distribution Function

- Distribution Function

  $F(y) = P(Y \le y)$



- Quantiles

  ❖ Quantiles of order τ:

  $F(y) = P(Y \le y) = \tau$ implies $y = F^{-1}(\tau) = Q(\tau)$

  Examples:

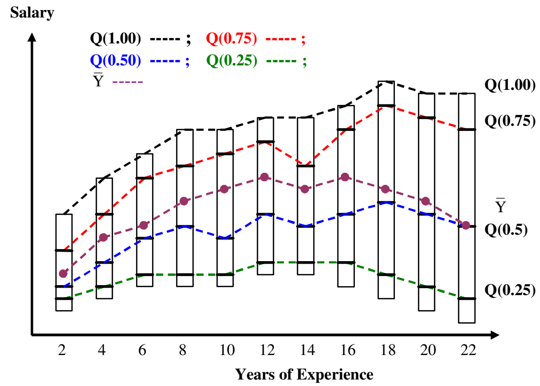
  $25 = F^{-1}(0.50) = Q(0.50)$

  $34 = F^{-1}(0.75) = Q(0.75)$

Quantile Regression

1. Classical Regression versus Quantile Regression

- Example: y = salary; x = years of experience



- Classical Regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

❖ Conditional Mean Regression

$$E(y_i \mid x_i) = \beta_0 + \beta_1 x_i \equiv \mu_i$$

- Quantile Regression of order $\tau \in [0, 1]$

$$y_i = \beta_0^\tau + \beta_1^\tau x_i + \varepsilon_i$$

❖ Conditional Quantile Regression

$$Q(\tau \mid x_i) = \beta_0^\tau + \beta_1^\tau x_i$$

2. Example:
   - Quantile Regressions of Wage Determination in Taiwan
   - Data: 29,133 plant workers in 1996

     wage = hourly wage;

     experience = months on current job
   - Regressions:

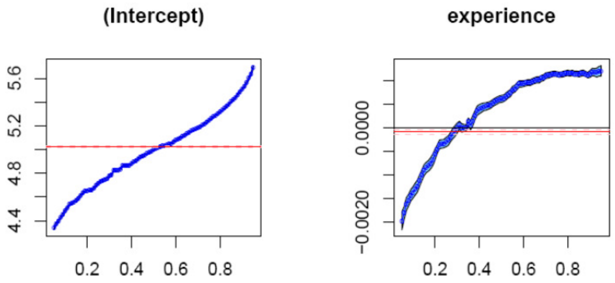
❖ Conditional Mean Regression

$$\ln(\text{wage}) = \beta_0 + \beta_1 (\text{experience}) + \varepsilon$$

❖ Quantile Regression

$$\ln(\text{wage}) = \beta_0^\tau + \beta_1^\tau (\text{experience}) + \varepsilon$$

τ	β_0^τ	β_1^τ
0.1	4.47987	-0.00137
0.2	4.65334	-0.00056
0.3	4.77952	0.00000
0.4	4.88014	0.00040
0.5	4.99839	0.00062
0.6	5.07736	0.00094
0.7	5.18110	0.00111
0.8	5.31419	0.00117
0.9	5.50766	0.00115
0.95	5.69128	0.00121
Conditional Mean	5.02400	-0.000086



Fundamentals of Quantile Regression Estimation

1. Estimation of Mean and Quantile

- Sample Observations: $y = (y_1, y_2, ..., y_n)$
- Population Mean and Population Quantile

  $E(y_i) = \mu; \quad Q(\tau) = F^{-1}(\tau)$
- Least-square Estimation of Population Mean

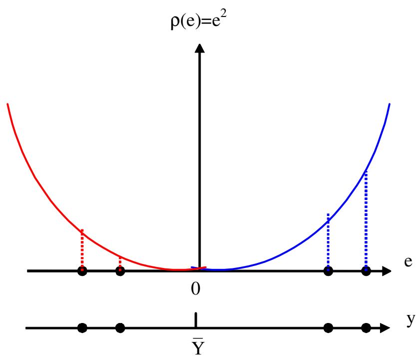
  ❖ Model: $y_i = \mu + \varepsilon_i = \bar{Y} + e_i$

  Error: $e_i = y_i - \bar{Y}$

❖ Loss Function: $\rho(e_i) = e_i^2 = (y_i - \bar{Y})^2$

❖ Least-square method:

$$\min_{\bar{Y}} \sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \bar{Y})^2; \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$



- Quantile Estimation of $Q(\tau)$
  - ❖ Model: $y_i = Q(\tau) + \varepsilon_i = \hat{Q} + e_i$

    Error: $e_i = y_i - \hat{Q}$

    Example: $\tau = 0.25$ implies

    $100\,\tau\,\%$ of $e_i = y_i - \hat{Q} \le 0$

    $100\,(1-\tau)\,\%$ of $e_i = y_i - \hat{Q} \ge 0$

  - ❖ Loss Function:

$$\rho_\tau(e_i) = \tau \times e_i = \tau \times \left(y_i - \hat{Q}\right) \quad \text{if } e_i \ge 0$$

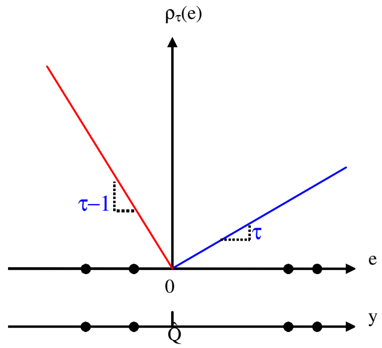
$$= (\tau - 1) \times e_i = (\tau - 1) \times \left(y_i - \hat{Q}\right) \quad \text{if } e_i \le 0$$

$$\rho_\tau(e_i) = \left(\tau - I(e_i < 0)\right) \times e_i$$

$$= \left(\tau - I\left(y_i - \hat{Q} < 0\right)\right) \times \left(y_i - Q\right)$$

❖ Quantile Estimation Method

$$\min_{\hat{Q}} \sum_{i=1}^{n} \rho_\tau(e_i) = \sum_{i=1}^{n} \left(\tau - I\left(y_i - \hat{Q} < 0\right)\right) \times \left(y_i - Q\right)$$



2. Estimation of Classical Regression and Quantile Regression

- Model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

- Classical Regression

  ❖ Conditional Mean: $E(y_i \mid x_i) = \beta_0 + \beta_1 x_i$

  ❖ Least-square Method:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2$$

- Quantile Regression:

  ❖ Conditional Quantile: $Q(\tau \mid x_i) = \beta_0^\tau + \beta_1^\tau x_i$

  ❖ Quantile Method:

$$\min_{\hat{\beta}_0^\tau, \hat{\beta}_1^\tau} \sum_{i=1}^{n} \rho_\tau(e_i) = \sum_{i=1}^{n} \left(\tau - I(e_i < 0) \right) \times (e_i)$$

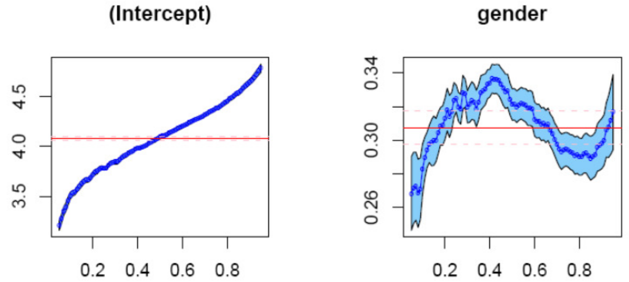
$$e_i = y_i - \hat{\beta}_0^\tau - \hat{\beta}_1^\tau x_i$$

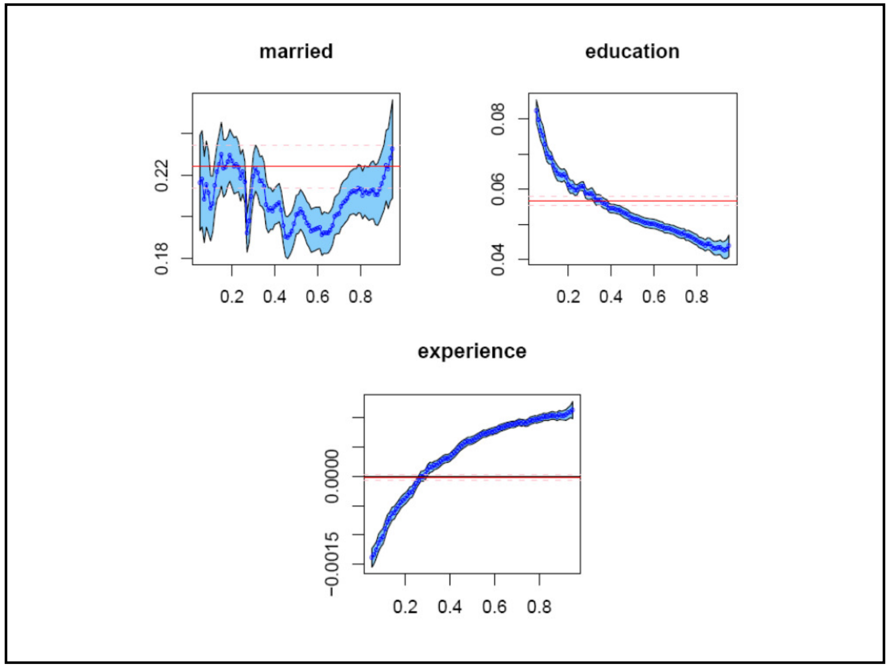
Computation of Quantile Regression: R-Program

- Free Download of R-Program at
  http://www.r-project.org

- Example on Wage Determination in Taiwan
  - Observations: 29,133 plant workers in 1996
  - Variables: wage = hourly wage; lnwage = ln(wage);
    gender = 1 male, 0 female; age = year of age;
    married = 1 married, 0 unmarried;
    education = year of education;
    experience = months on current job

(3) Graphs of β^{τ}




married
0.18
0.22
0.2
0.4
0.6
0.8
education
0.04
0.06
0.08
0.2
0.4
0.6
0.8
experience
−0.0015
0.0000
0.2
0.4
0.6
0.8