Lab 6: Unconditional Quantile Regression

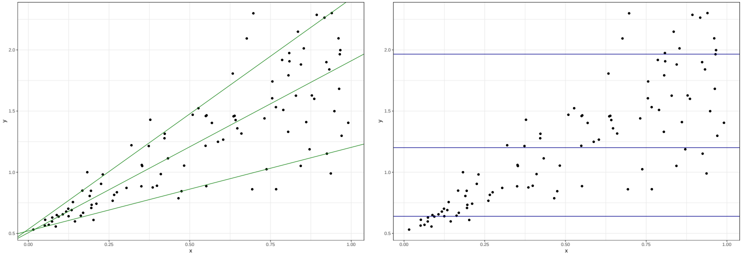
Econometrics Beyond Ordinary Least Squares

Patrick Mokre

WS 2020/2021

Unconditional Quantile Regression

Unconditional Quantiles: Quantiles of the overall distribution.



(a) Conditional Quantiles

(b) Unconditional Quantiles

Figure 1: Conditional and Unconditional Quantiles

Conditional and Unconditional Quantile Regression Coefficients

CQR coefficients: Workers' benefit of **transferring into** public sector jobs.

UQR coefficients: Earnings increase if **percentage of public sector workers** increases.

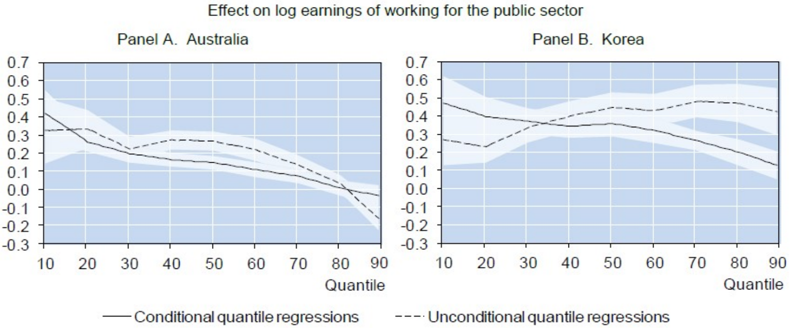


Figure 2: Fournier, 2012, Fig 3: "Conditional and unconditional quantile regression estimates of the impact on earnings from working in the public sector"

UQR and Inequality

Changes in unconditional quantiles $\Rightarrow$ retrieve **changes in quantile ratios** (eg. 80/20, 90/10).

Extend UQR to other distributional properties $\Rightarrow$ retrieve changes in **Gini coefficient, mean-median ratio**.

In OLS $\hat{\beta} = E(Y \mid X = 1) - E(Y \mid X = 0)$, the conditional effect of transferring between groups, as well as $\hat{\beta} = \frac{\partial \mu(p)}{\partial p}$ with $p = \frac{\sum_i^N 1(X=1)}{\sum_i^N 1(X=1) + 1(X=0)}$.

This **does not hold for CQR**, usually $\hat{\beta}_\tau = F_Y^{-1}(\tau \mid X = 1) - F_Y^{-1}(\tau \mid X = 0) \neq \frac{\partial q_\tau(p)}{\partial p} = P(Y > q_\tau \mid X = 1) - P(Y > q_\tau \mid X = 0)$.

Influence Functions and Recentered Influence Functions

"*The influence function $IF(Y; v, F_Y)$ of a distributional statistic $v(F_Y)$ represents the influence of an individual observation on that distributional statistic.*" (Firpo et.al., 2009, 954)

Re-centered influence function RIF: $RIF(Y; v, F_Y) = v(F_Y) + IF(Y; v, F_Y)$.

Note: $E[RIF(y; v, F_Y)] = E[v(F_Y)]$.

$$IF(Y; q_\tau, F_Y) = \frac{\tau - 1(Y \leq q_\tau)}{f_Y(q_\tau)} \tag{1}$$

$$RIF(Y; q_\tau, F_Y) = q_\tau + IF(Y; q_\tau, F_Y) \tag{2}$$

RIF Regressions

RIF regression model: conditional expectation of the RIF $E[RIF(Y; q_\tau, F_Y) \mid X]]$.

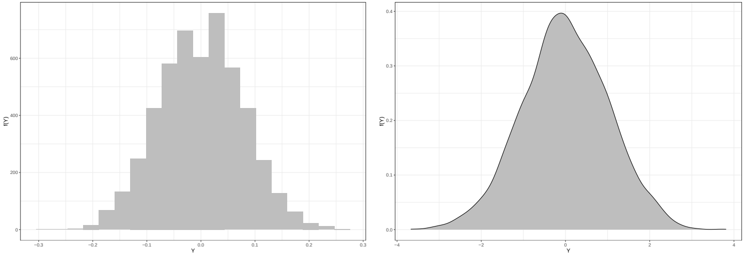
RIF regression: OLS estimation $RIF(Y; q_\tau, F_Y) \mid X_i] = \beta_{RIF} X_i + \epsilon_i$.

$\hat{\beta}_{RIF}$ corresponds to the **effect of a marginal change in** X on the unconditional quantile of Y.

Necessary steps: Estimate quantiles $q_\tau(Y)$, density $f_Y(q_\tau)$ (eg. by Kernel estimation), calculate dummy variable $1(Y \leq q_\tau)$ (trivial).

Distributions: Probability Density Function f(Y)

For a stochastic experiment with outcome variable Y, the **probability density function (PDF)** gives the probability of a certain realization $Y = y$ to be observed.



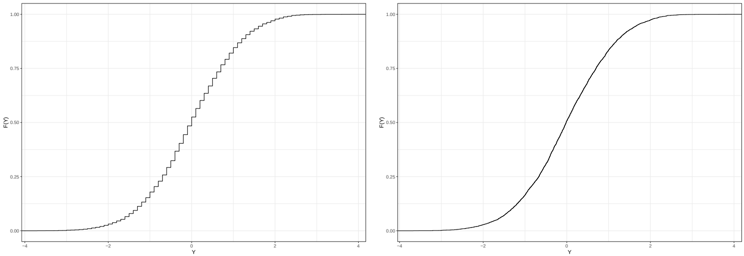
(a) Dscrete (b) Continuous

Figure 3: Probability Density Functions of a Discrete and a Continuous Variable Y N(0,1)

Distributions: Cumulative Density Function F(Y)

The cumulative density function (CDF) gives the probability that a random variable **realizes below a treshold level** $P(Y < y)$. It is 0 for the minimum and 1 for the maximum range of Y.

For observations it can be understood as the fraction of the population with realizations below some observation.



(a) Dscrete

(b) Continuous

Figure 4: Cumulative Density Functions of a Discrete and a Continuous Variable Y N(0,1)

Joint Distributions

The **joint distribution** $f_{X,Y}(x, y)$ of two variables X, Y gives the probability of observing two values $X = x$, $Y = y$ **at the same time**.

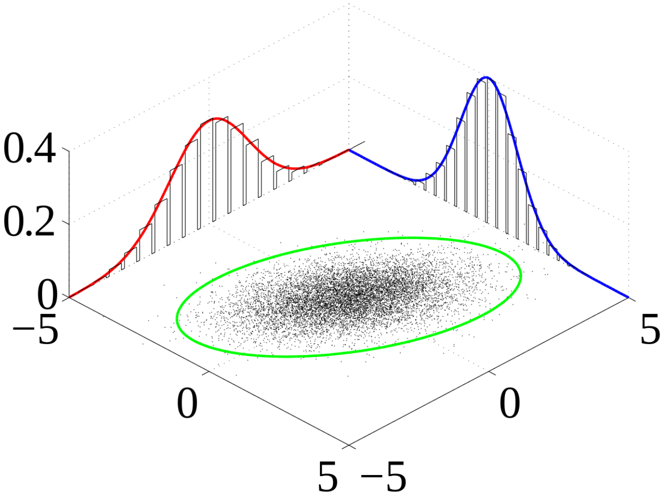


Figure 5: Wikipedia: "Many sample observations (black) are shown from a joint probability distribution. The marginal densities are shown as well."

Marginal Densities

Influence Functions

A **functional** $v(F_Y)$ projects from a function F_Y to the space of real numbers $\mathbb{R}$, ie. $v : F_v \rightarrow \mathbb{R}$, eg. **the mean**.

Hampel, 1968: The influence function gives the infinitesimal behavior of a functional v.

One can have two distributions of the same class F_Y and G_Y (eg. one $N(0,1)$ and one $N(1,1)$).

Then there exists a **mixing distribution that is t units away** from F_Y in the direction of G_Y: $F_{Y,t \cdot G_Y} = (1-t)F_Y + tG_Y = t(G_Y - F_Y) + F_Y$.

$$\lim_{t \downarrow 0} \frac{v(F_{Y,t \cdot G_Y}) - v(F_Y)}{t} = \frac{\partial v(F_{Y,t \cdot G_Y})}{\partial t}\Big|_{t=0}$$
$$= \int IF(y; v, F_Y) \cdot d(G_Y - F_Y)(y) \quad (3)$$

Recentered Influence Function

The *von-Misès linear approximation* of $v(F_{Y,t\cdot G_Y}) - v(F_Y)$:

$$v(F_{Y,t\cdot G_Y}) - v(F_Y) = v(F_Y) + \int IF(y; v, F_Y) \cdot f(G_Y - F_Y)(y) \\ + r(t; v; G_Y, F_Y) \quad (4)$$

Neutralize the "remainder term" $r(.)$ (by setting $G_Y = \Delta_y$ and $t = 1$), for the **re-centered influence function RIF**

$$RIF(y; v, F_Y) = v(F_Y) + \int IF(s; v, F_Y) d\Delta_y(s) = v(F_Y) + IF(y; v, F_Y) \quad (5)$$

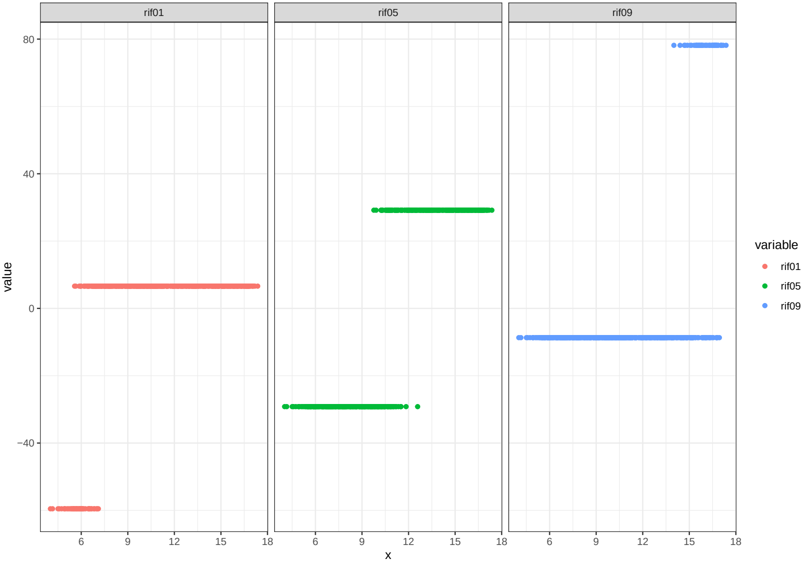
Key result: "*the impact of a marginal change in the distribution of X on $v(F_Y)$ can be obtained using the conditional expectation of the $RIF(Y; v, F_Y)$*". (Firpo et.al., 2009, 957)

Calculating the RIF

$$RIF(Y; q_\tau, F_Y) = q_\tau + IF(Y; q_\tau, F_Y)$$
$$= q_\tau + \frac{\tau - 1(Y \leq q_\tau)}{f_Y(q_\tau)} \quad (6)$$

```
set.seed(1)
edu_m <- runif(500, 4, 16) + rnorm(500, 1, 0.5)
inc_m <- rnorm(500, 5*edu_m, 1 + 0.3*edu_m)
af_m <- approxfun(density(inc_m))
df_tmp <- data.frame(x = edu_m,
                     y = inc_m) %>%
  mutate(rif01 = (0.1 - ifelse(y>quantile(inc_m, 0.1), 0, 1))/
           af_m(quantile(inc_m, 0.1)),
         rif05 = (0.5 - ifelse(y>quantile(inc_m, 0.5), 0, 1))/
           af_m(quantile(inc_m, 0.5)),
         rif09 = (0.9 - ifelse(y>quantile(inc_m, 0.9), 0, 1))/
           af_m(quantile(inc_m, 0.9)))
```

Bivariate RIF



RIF: OLS estimation

$$RIF(y_i; q_\tau, F_Y) \sim \alpha + X_i'\beta + \epsilon_i \tag{7}$$

$$\beta = \frac{\partial RIF(y_i; q_\tau, F_Y)}{\partial X_i} \tag{8}$$

```
df_tmp %>%
  lm(rif01 ~ x, data=.)
```

Call: lm(formula = rif01 ~ x, data = .)

Coefficients: (Intercept) x
-31.718 2.902

RIF: Coefficient Decomposition

- Same regression between two groups: different **endowments** and **coefficients** (eg. women have more years of education, but smaller income returns for each year).
- *Oaxaca (1973)* and *Blinder (1973)*: Decompose effects by calculating **counterfactual distribution if group A had endowments of group B**.

$$\Delta\bar{Y} = \bar{Y}_A - \bar{Y}_B$$
$$\Delta\bar{Y} = \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B$$
$$\Delta\bar{Y} = (\bar{X}_A - \bar{X}_B)'\hat{\beta}_B + \bar{X}'_B(\hat{\beta}_A - \hat{\beta}_B) + (\bar{X}_A - \bar{X}_B)'(\hat{\beta}_A - \hat{\beta}_B) \tag{9}$$

- UQR estimated by OLS: Oaxaca-Blinder decomposition is trivial.
- Interpretation of Coefficients is not.
- In surveys, re-weighting becomes crucial.