Topics in Applied Econometrics for Public Policy

Master in Economics of Public Policy, BSE

Handout 6: Quantile Regression, II

Laura Mayoral

IAE and BSE

Barcelona, Spring 2024

Quantile Regression: Summary (so far)

- In any empirical analysis relating Y and X, we can be interested in several aspects of the condional distribution $Y|X = x$

- Values that estimate the central tendency of this distribution: conditional mean, conditional median,

- Values that look at the dispersion of the distribution: conditional quantiles (that look at aspects other than the median),

- QR is typically employed with continuous dependent variables (so the quantiles are uniquely defined), but there are exceptions (e.g., count data)

- QR estimators can be obtained by optimizing an objective function (average of the check function. $\rho(.)$), in a similar way as we do when we do develop OLS estimators.

$$\hat{\beta}_\tau = \underset{b}{\operatorname{argmin}} \sum_{i=1} \rho_\tau(|Y_i - X_i' b|)$$

- There's no close-form solution (unlike in OLS), optimization is done numerically

- Special case: LAD (least absolute deviations).
  - Estimates the conditional median
  - Advantanges and disadvantages w.r.t. OLS.
  - A good option if the data contains outliers

- Estimation is easy, interpretation has to be done with care

This handout: Roadmap

1. Interpretation of coefficients (cont.)

2. Asymptotic properties

3. Estimation of Standard Errors. Confidence Intervals

4. QR with Panel Data

5. QR with censored data

6. Non parametric quantile regression

7. Quantile causal effects

More on interpretation: Retransformation

- In the example of the previous handout: dependent variable is log expenditures.

- We're interpreting the impact of variable X on logY in quantile τ, are we interested on this?

- We're typically interested in the effect of X on Y (not on logY).

- Question: if we've estimated a QR where the dependent variable is g(Y) (g is monotonic and increasing function) then how do we interpret marginal effects with respect to Y?

- Before we answer this question, let's consider first transformations of a variable, its quantiles and expectations.

- Consider a variable Z, g(Z), where g is a monotonic transformation. For instance $g(Z) = Z^2, Z > 0$

  - If you know that E(g(Z))=6, what's the value of E(Z)?

- Before we answer this question, let's consider first transformations of a variable, its quantiles and expectations.

- Consider a variable Z, g(Z), where g is a monotonic transformation. For instance $g(Z) = Z^2, Z > 0$

  - If you know that E(g(Z))=6, what's the value of E(Z)?

  - If you know that the 40th percentile of g(Z) is 4, what's the value of the 40th percentile of Z?

- As you can see, expectations and quantiles behave differently when transformations are made. We need to take this into account when interpreting marginal effects.

- Equivalence property of QR: Given $q_\tau(g(Y)|Z) = X'\beta$, (g is monotonic and invertible) then $q_\tau(Y|Z) = g^{-1}(X'\beta)$

- For example: $q_\tau(lnY|Z) = X'\beta \Rightarrow q_\tau(Y|X) = e^{(X'\beta)}$

- Let's go back to the computation of marginal effects.

- Let's derive $q_\tau(Y|X)$ with respect to X_j:

$$\frac{\partial q_\tau(Y|X)}{\partial X_j} = \frac{\partial e^{(X'\beta)}}{\partial X_j} = e^{(X'\beta_\tau)}\beta_{\tau j}$$

- The derivative depends on X.

- Average marginal effect (AME):

$$N^{-1}\sum_{i=1}^{N} e^{(X_i'\beta_\tau)}\beta_{\tau j}$$

Using STATA:

qreg ltotexp totchr age female white

quietly predict xb

gen expxb=exp(xb)

quietly sum expxb

display "Multiplier of QR in logs coeffs to get AME in levels =" r(mean)

```
.  regress ltotexp totchr age female white

      Source |       SS           df       MS      Number of obs   =     2,955
-------------+----------------------------------   F(4, 2950)      =    171.39
       Model |  1041.82144         4  260.455359   Prob > F        =    0.0000
    Residual |  4483.06795     2,950  1.51968405   R-squared       =    0.1886
-------------+----------------------------------   Adj R-squared   =    0.1875
       Total |  5524.88938     2,954  1.87030785   Root MSE        =    1.2328

------------------------------------------------------------------------------
     ltotexp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      totchr |   .4476954   .0176308    25.39   0.000     .4131256    .4822653
         age |   .0102383   .0035862     2.85   0.004     .0032067      .01727
      female |  -.0952113   .0462155    -2.06   0.039    -.1858292   -.0045934
       white |   .3582365   .1418762     2.52   0.012       .08005    .6364229
       _cons |   6.196758   .2921724    21.21   0.000     5.623876    6.769641
------------------------------------------------------------------------------

.
.  quietly predict xb

.
. gen expxb=exp(xb)

.
. quietly sum expxb

.
. display "Multiplier of QR in logs coeffs to get AME in levels =" r(mean)
Multiplier of QR in logs coeffs to get AME in levels =3761.9663
```

- To compute marginal effects of X_j on Y, just multiply 3761 by the relevant coefficient $\beta_{j\tau}$

- Final Note: the equivalence property of QR is exact only if the conditional quantile function is correctly specified.

- In applications this is not generally the case, so it has to be interpreted as an approximation

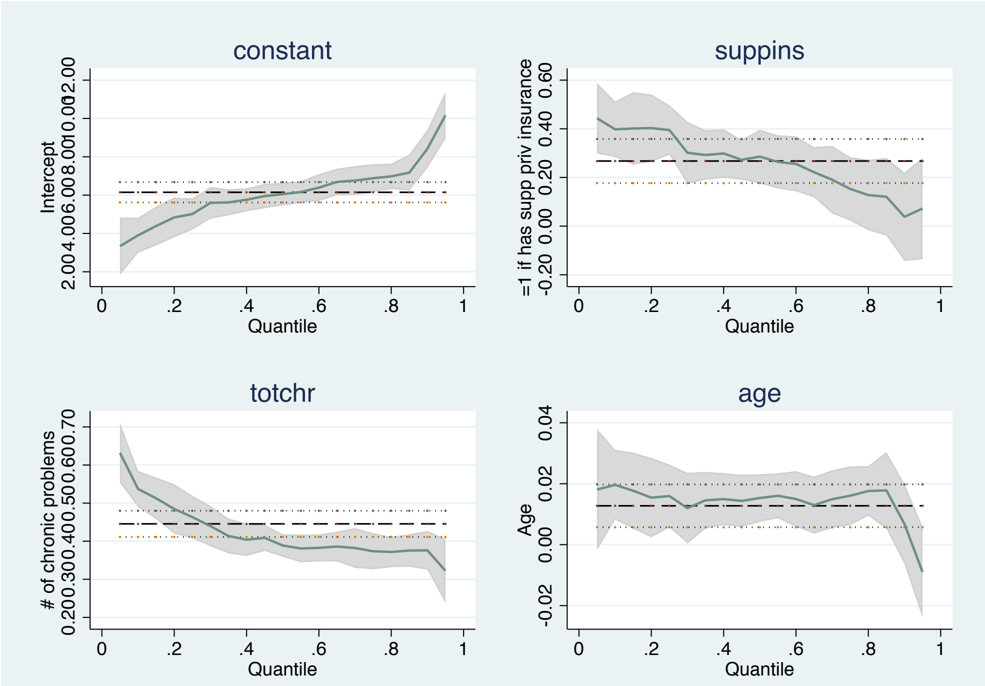
Graphical display of coefficients over quantiles

- When we estimate QR for different values of τ there are a lot of coefficients to analyze

- Graphical representations of the results are very useful

- One possibility is to construct one graph for each variable in the regression that displays how β_τ changes for $\tau \in (0,1)$

- Horizontal line: OLS point estimates and CI (constant across quantiles)



- In STATA command: grqreg
- You need to install the command first
- the code used to generate the previous graph:
- The graph includes the ols coefficients
- ci and ciols include the confidence interval for the ols and QR coefficients

```
ssc install grqreg

bsqreg ltotexp suppins totchr age , reps(100)

grqreg, cons ci ols olsci title(constant suppins totchr age)
```

2. Asymptotic Properties of the QR estimator

- Model: The linear quantile regression model is

$$Y = X'\beta_\tau + e$$

$$q_\tau(e|X) = 0$$

- These two equations imply that the conditional quantile τ of Y given X is $X'\beta_\tau$

- Notice that the error e is not centered at zero, instead it's centered so that its τth quantile is zero.

- This is a normalization, but it changes the role of intercept changes when we move from mean regression to QR.

- Recall that (the population) β_τ can be obtained as

$$\beta_\tau = argmin_b E[\rho_\tau(Y - X'b)].$$

- The QR estimator of β_τ, $\hat{\beta}_\tau$ is given by the sample analog of β_τ:

$$\hat{\beta}_\tau = argmin_b \frac{1}{N} \sum_{i=1}^{N} \rho_\tau(Y_i - X_i'b)$$

Consistency

- Under broad (and a bit technical) assumptions, the QR estimator is consistent:

$$\hat{\beta}_{\tau} \xrightarrow{p} \beta_{\tau}$$

(From Hansen's book):

Theorem 24.3 Consistency of Quantile Regression Estimator
Assume that (Y_i, X_i) are i.i.d., $\mathbb{E}|Y| < \infty$, $\mathbb{E}\left[\|X\|^2\right] < \infty$, $f_\tau(e \mid x)$ exists and satisfies $f_\tau(e \mid x) \le D < \infty$, and the parameter space for β is compact. For any $\tau \in (0,1)$ such that

$$\boldsymbol{Q}_\tau \overset{\text{def}}{=} \mathbb{E}\left[XX' f_\tau(0 \mid X)\right] > 0 \qquad (24.18)$$

then $\widehat{\beta}_\tau \underset{p}{\longrightarrow} \beta_\tau$ as $n \to \infty$.

- Technical note:

- Condition 24.18 is needed for the uniqueness of the coefficients β_τ

- A sufficient condition (also called quantile independence): assume that the cond. distribution of the error e doesn't depend on X at $e = 0$, thus 24.18 simplifies to

$$Q_\tau = E(XX')f_\tau(0)$$

- Advice: there's no need to assume this (this is a strong assumption)

- The reason we highlight this is because STATA's default uses this suficient condition to compute the var-cov matrix of $\hat{\beta}_\tau$(we'll see that in a few slides).

Asymptotic Normality

- $\hat{\beta}_\tau$ is $\sqrt{N}$-consistent and asymptotically normal

Theorem 24.4 Asymptotic Distribution of Quantile Regression Estimator
In addition to the assumptions of Theorem 24.3, assume that $f_\tau(e \mid x)$ is continuous in e, and β_τ is in the interior of the parameter space. Then as $n \to \infty$

$$\sqrt{n}\left(\widehat{\beta}_\tau - \beta_\tau\right) \underset{d}{\longrightarrow} \mathrm{N}\left(0, \boldsymbol{V}_\tau\right)$$

where $\boldsymbol{V}_\tau = \boldsymbol{Q}_\tau^{-1}\Omega_\tau\boldsymbol{Q}_\tau^{-1}$ and $\Omega_\tau = \mathbb{E}\left[XX'\psi_\tau^2\right]$ for $\psi_\tau = \tau - \mathbb{1}\left\{Y < X'\beta_\tau\right\}$.

And if the model is not exactly linear?

- Assume that the conditional quantile τ is not linear but a linear model is estimated.

- What is the linear model estimating in this case?

- Remember: when the same happens in an OLS framework, the linear model is the best linear approximation to the conditional expectation.

- Luckily, the same happens in the case of quantile regression, a linear model can be also interpreted as the best linear approximation the conditional quantile.

- See Mostly Harmless p.277 for additional details.

- Therefore:

- The results above don't assume correct specification

- This means that we can interpret the linear function as an approximation to the "true function", we don't need the "truth" to be exactly linear

- Then: the variance-covariance matrix in theorem 24.4. is the most general and applies broadly for practical applications where linear models are approximations (rather than literal truths)

- This variance-covariance matrix simplifies if we impose different assumptions, for instance
  - correct specification
  - quantile independence

These are the expressions of the var-cov matrix under different assumptions (from Hansen's book):

Combined with (24.19) we have three levels of asymptotic covariance matrices.

1. General: $\boldsymbol{V}_\tau = \boldsymbol{Q}_\tau^{-1}\Omega_\tau \boldsymbol{Q}_\tau^{-1}$
2. Correct Specification: $\boldsymbol{V}_\tau^c = \tau(1-\tau)\boldsymbol{Q}_\tau^{-1}\boldsymbol{Q}\boldsymbol{Q}_\tau^{-1}$
3. Quantile Independence: $\boldsymbol{V}_\tau^0 = \dfrac{\tau(1-\tau)}{f_\tau(0)^2}\boldsymbol{Q}^{-1}$

Advice: Always take as few assumptions as possible. Therefore, go for the first expression, as it's valid under broad conditions unlike the other two!

3. Estimation of the Variance-Covariance matrix: some tips

- By default, STATA qreg doesn't estimate V_τ (the general variance-covariance matrix that allows for mispecification and is derived under general conditions)

- Instead, it provides standard errors based on V_τ^0, the var-cov matrix under correct specification and quantile independence (see Hansen).

- You should avoid the use of these standard errors (for identical reasons you should avoid homocedastic variance-covariance matrices in OLS).

- If you use vce(robust): variance-covariance matrix that still assumes correct specification but drops the quantile independence assumption (V_τ^c).

- For a more general variance-covariance matrix estimate: use Bootstrap

Estimation of the Variance-Covariance matrix: some tips, II

- Some tips to compute Bootstrap std. errors for QR regression

- STATA command:

bootstrap, reps(#): qreg y x

- Number of replications should be large: at least 1000 (10,000 preferred!)

- Time consuming, only needs to be done for your final calculations (i.e., do intermediate regressions with less replications to save time).

Bootstrap confidence intervals

- Two ways of computing Bootstrap CI

  a) Use bootstrap std.error and Normal quantile

  b) Use percentiles of bootstrap distribution

- For obvious reasons the second way is better! but this is not the STATA default

- To obtain b) type:

```
bootstrap, reps(#): qreg y x

estat bootstrap
```

Clustered Standard Errors

- Not implemented by qreg

- Can be obtained in the bootstrap case:

  bootstrap, reps(#) cluster(id): qreg y x

  estat bootstrap.

- STATA tip:

- Some of the QR built-in commands can be very slow, particularly when bootstrap std. errors are computed.

- Alternative user-written package: IVQTE (Blaise Melly),

see here

Takeaways

- QR estimator is consistent and asymptotically normal under broad assumptions (it doesn't require correct specification)

- Use std. errors valid under broad assumptions

- In practice: use bootstrap standard errors

- To compute CI: use percentiles from the bootstrap distribution

4. Panel data

- Assume now we have panel data: $\{Y_{it}, X_{it}\}, i = 1, ..., N, t = 1, \ldots, T$.

- A natural model to consider: a linear model with an individual effect $\alpha_{i\tau}$

$$Q_\tau[Y_{it} \mid X_{it}, \alpha_{i\tau}] = X_i' \beta_\tau + \alpha_{i\tau}.$$

- Can we apply any of the techniques we typically employ in standard panel regressions to get rid of α_i?

- Recall that these methods are:

1. Remove the individual effect by the within transformation (i.e., for each individual, subtract its mean, see section 17.8 in Hansen's book for details);

2. Remove the individual effect by first differencing;

3. Estimate a full quantile regression model using the dummy variable representation.

Panel data, II

- Unfortunately, all of these methods fail!
- Why?

- Methods (1) and (2) fail for the same reason: The quantile operator Q_τ is not a linear operator:

  - The within transformation of $Q_\tau[Y_{it}|X_{it}, \alpha_{i\tau}]$ does not equal $Q_\tau[\tilde{Y}_{it}|X_{it}, \alpha_{i\tau}]$

  - similarly $\Delta Q_\tau[Y_{it}|X_{it}, \alpha_{i\tau}] \neq Q_\tau[\Delta Y_{it}|X_{it}, \alpha_{i\tau}]$.

- Method (3) fails because of the incidental parameters problem:

  - the number of parameters in the model (because of the individual dummies) is proportional to sample size

  - in this context, nonlinear estimators (including quantile regression) are inconsistent.

Panel data, III

- QR estimators for Panel data: several proposals to deal with this issue, but none are particularly satisfactory.

- Canay (2011)'s method: has the advantage of simplicity and wide applicability.

- Based on a simplification: the individual effect is common across quantiles: $\alpha_{i\tau} = \alpha_i$.

- Thus α_i shifts the quantile regressions up and down uniformly.

- Under this assumption: we can write the quantile regression model as

$$Y_{it} = X_i' \beta_\tau + \alpha_i + e_{it}$$

Panel data, IV

- Canay's estimator takes the following steps:

1. Estimate α_i by (standard) fixed effects $\hat{\alpha}_i$
2. Estimate $\beta(\tau)$ by quantile regression of $Y_{it} - \hat{\alpha}_i$ on X_{it}.

- How to do step 1:

- The key: the assumption that the fixed effect α_i does not vary across the quantiles τ, means that the fixed effects can be estimated by conventional fixed effects.

- Then, use a fixed effect estimator for the conditional mean. The model for the conditional mean would be

$$Y_{it} = X_i'\theta + \alpha_i + e_{it}$$

- more specifically: Estimate θ by the within estimator and α_i by taking averages of $Y_{it} - X_{it}\hat{\theta}$.

- How to do step 2:

- After step 1, estimate $\beta(\tau)$ by quantile regression of $Y_{it} - \hat{\alpha}_i$ on X_{it}.

- Primary disadvantage of this approach: the assumption that α_i does not vary across quantiles is restrictive.

- This is a topic of active research

- More contributions: visit Blaise Melly website for recent contributions (with STATA packages)

Melly and Pons (2023).

5. Censored data and QR

- Censored data: a situation in which not all the values of the distribution are provided, typically very small/large values are given in an interval

- Example: wages. Frequently, high wages are grouped in one category, i.e., wages$> 100,000$ a year (top-coded)

- Estimators of the mean (conditional mean) are not consistent: we need all the distribution in order to compute expectations correctly

- However, this problem doesn't affect quantiles! In this example: all quantiles below the censoring point are unaffected by the censoring.

- More formally:

- if the variable y is top-coded above a value c, we observe $Y^* = \min(y, c)$ instead of y.

- Then, using an idea by Powell (1986), we can exploit the fact that $q_\tau(Y^*|X_i) = \min(X_i'\beta_\tau, c)$.

- The parameter vector β_τ^c solves

$$\hat{\beta}_{\tau,c} = \underset{b}{argmin} E(1[X'b < c](\rho_\tau(Y - X'b))). \tag{1}$$

Hence, we estimate β_τ as:

$$\hat{\beta}_{\tau,c} = \arg\min_b \sum_{i=1}^{N} \left(1[X_i'b < c]\ \rho_\tau(Y_i - X_i'b)\ \right). \tag{2}$$

6. Non parametric quantile regression

- So far, we've assumed that quantile regression functions are linear

- We know that this is a simplification, in many instances, an over simplification.

- Good news: Quantile regression functions may be estimated using standard nonparametric methods.

- This is a potentially large subject.

- The simplest way to go: consider series methods, which have the advantage that they are easily implemented with conventional software.

Non parametric quantile regression, II

- The nonparametric quantile regression model is

$$Y = g_\tau(X) + e$$

$$Q_\tau[e|X] = 0$$

- Idea of series regression: approximate the function $g\tau(.)$ by a series regression as the ones we saw in Handout 4.

- For example,

$$Y = \beta_0 + \beta_1 X + \cdots + \beta_k X^k + e_k$$

with $Q_\tau[e_k|X] \approx 0$

- For any k, the coefficients can be estimated by quantile regression.

- As in series regression the model order k should be selected to trade off flexibility (bias reduction) and parsimony (variance reduction).

- Caveat: how to select k in a given application?

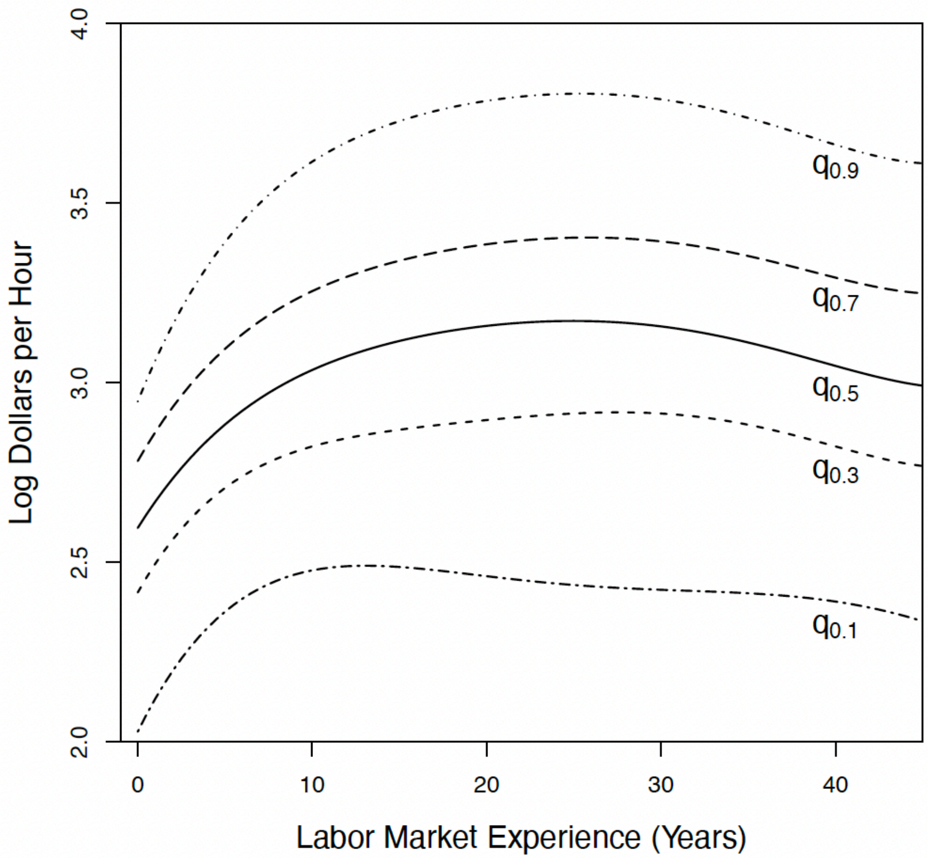
- Unfortunately, standard information criterion (such as the AIC) do not apply for quantile regression

- It is unclear if crossvalidation is an appropriate model selection technique.

- These questions are an important topic for future study.

An example (Hansen, p.796)

Y: log wage quantile regressions on a 5th order polynomial in experience.



- There are two notable features.

- First, the $\tau = .1$ quantile function peaks at a low level of experience (about 10 years) and then declines substantially with experience.

- Second, even though this is in a logarithmic scale the gaps betwen the quantile functions substantially widen with experience. This means that heterogeneity in wages increases more than proportionately as experience increases.

7. Causality: Quantile Causal Effects

- Key question: Can we interpret the results obtained in QR as causal?

- We can partially answer this question in the treatment response framework

- We will provide conditions under which the quantile regression derivatives equal quantile treatment effects.

Treatment-Response model

- Y is outcome, X are controls and D is the treatment variable, U is an unobserved structural random error.

- For concreteness: Y: wage, D: college education; U: (unobserved) ability

$$Y = h(D, X, U)$$

- For simplicity, D is binary: $D = 0$ or 1.

- Causal effect of the treatment

$$C(X, U) = h(1, X, U) - h(0, X, U).$$

- In general, this effect is heterogeneous across individuals: we can study different aspects of its distribution, in particular, mean and quantiles

- Average Treatment Effect: average of heterogeneous treatment effect, $E[C(X,U)|X=x]$

- Quantile treatment effect is its τth conditional quantile

$$q^*_\tau(x) = q_\tau[C(X,U)|X=x].$$

- Interpretation of $q^*_\tau(x)$: traces out the distribution of the causal effect across the different quantiles

- Notice that it looks at the quantiles of the distribution of the causal effect (=the difference between the cases D=1/D=0)

- From observational data, we can estimate the quantile regression function (as we've done up to now)

$$q_\tau(d,x) = q_\tau[Y|D=d, X=x] = q_\tau[h(D,X,U)|D=d, X=x]$$

- The estimated effect of D would be

$$D_\tau(X) = q_\tau(1,x) - q_\tau(0,x)$$

- Key Question: Under what conditions $D_\tau(X) = q^*_\tau(x)$

- Notice the difference:

- $D_\tau(X)$: difference of conditional quantiles (τ) of people with college and people without college

- $q^*_\tau(x)$: quantile τ of the effect of going to college.

- The required conditions are (see Hansen p. 793):

Assumption 24.1 Conditions for Quantile Causal Effect

1. The error U is real valued.
2. The causal effect $C(x, u)$ is monotonically increasing in u.
3. The treatment response $h(D, X, u)$ is monotonically increasing in u.
4. Conditional on X the random variables D and U are independent.

- Theorem:

Under Assumption 24.1, $D_\tau(X) = q_\tau^*(x)$

- To understand the theorem we need to understand the meaning of these conditions, let's consider an example:

- Example: impact of college attendance on wages;

Y: wages,

D: college attendance;

U: innate ability (unobserved, not in the model).

X: a bunch of control variables

Meaning of assumptions in 24.1

- Assumption 24.1.1: excludes multi-dimensional unobserved heterogeneity.

- Assumption 24.1.2 & Assumption 24.1.2: monotonicity assumptions

  - Assumption 24.1.2 requires that the wage gain from attending college is increasing in latent ability U (given X).

  - Assumption 24.1.2 requires that wages are increasing in latent ability U whether or not an individual attends college.

- To see the role of these two assumptions, consider two individuals A and B, A has higher ability than B. These two assumptions together require

  - A's gain from attending college exceeds B's gain.

  - A receives a higher wage than B if they both are high school graduates AND if they are both college graduates

More on assumptions

- Assumption 24.1.4 is the traditional conditional independence assumption.

- This is a critical condition for causal inference:

By conditioning on a sufficiently rich set of variables X any correlation between D and U has been eliminated.

- Under this condition, the probability of receiving the treatment (conditioning on observables) doesn't depend on unobserved variables.

$$P(D = 1|X, U) = P(D = 1|X)$$

- (But notice how stringent this assumption is, under this assumption, the probability of attending college doesn't depend on ability!)

- It's clear that these conditions won't hold in many applications.

- Solution: instrumental variables

Takeaways

- Under the conditional independence and the monotonicity assumptions, the quantile regression coefficients are the marginal causal effect of the treatment variable D upon the distribution of Y

- The coefficients are not the marginal causal effects for specific individuals, rather they are the causal effect for the distribution.

- As in the conditional mean case, these conditions can be very demanding

- For instance, in the example above, is it reasonable to expect that attending college is unrelated to (unobserved) innate ability?

- What if they don't hold?

The IV QR

- As in the OLS case, endogeneity can be solved by using a good instrument(s)

- Same idea: the instrument should verify an uncorrelatedness/independend assumption

- IV methods for quantile regressions, however, are not so simple, and are still under development these days.

- We'll focus on a particular case: estimation of treatment effects

IV estimation of Quantile Treatment Effects (QTE)

- A particular case:
  - D (treatment) is binary and Z (instrument) is binary

- Under these assumptions, Abadie, Angrist and Imbens (2002) introduced an IV estimator that is simple to implement:

- Quantile treatment effect estimator

- Their paper: "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings", Econometrica 2002.

Quantile Treatment Effects Estimator: Framework

- Similar assumptions as LATE framework for average causal effects.
- LATE: local average treatment effect

- Setup:
- Binary treatment D
- Potential endogeneity due to omitted variables
- A binary instrument Z is available
- We can think of Z as initiating a causal chain: Z$\Rightarrow D \Rightarrow Y$

An example

- Question: do the poorest workers benefit from a training program?

- Binary treatment: doing the training program or not.

- How is the treatment assigned: lottery

- However, participation is voluntary, so workers self-select themselves to treatment.

- (Binary) instrument: being assigned to treatment by the lottery (intention to treat).

- To capture the idea that Z has a causal effect on D consider this notation:

$$D_{1i}\text{: i's treatment status if } Z_i = 1$$

and

$$D_{0i}\text{: i's treatment status if } Z_i = 0$$

- The LATE framework partitions any population with an instrument into three sets of instrument-dependent subpopulations.
  - Compliers: $D_{1i} = 1$ and $D_{0i} = 0$
  - Always takers: $D_{1i} = 1$ and $D_{0i} = 1$
  - Never takers: $D_{1i} = 0$ and $D_{0i} = 0$

- To capture the idea that Z has a causal effect on D consider this notation:

$$D_{1i}\text{: i's treatment status if } Z_i = 1$$

and

$$D_{0i}\text{: i's treatment status if } Z_i = 0$$

- The LATE framework partitions any population with an instrument into three sets of instrument-dependent subpopulations.
  - Compliers: $D_{1i} = 1$ and $D_{0i} = 0$
  - Always takers: $D_{1i} = 1$ and $D_{0i} = 1$
  - Never takers: $D_{1i} = 0$ and $D_{0i} = 0$
- (How about the "defiers"? Monotonity assumption: $P(D_1 \geq D_0|X) = 1$.)

- The “local” nature of LATE:

- We can only identify the effect of the treatment on the population of compliers. Why?

- The instrument is not informative in the population of always takers or never takers

- For these groups, the instrument is not a “source of exogeneous variation”, as by definition treatment status for these two groups is unchanged by the instrument.

- The effect in the whole population might be different than the “local” effect

Example:

- Goal: estimating the effect of attending college on wages at different points in the distribution.

- Problem: the decision of attending college is not random, potentially depends on unobserved variables (e.g., ability)

- We need an instrument that generates a group of compliers: give a (random) subsidy

  - Compliers: those that attend college with the subsidy but wouldn't do it without it.

  - Always takers: always go to college, regardless of the subsidy

  - Never takers: never go to college, regardless of the subsidy

- For obvious reasons, the use of the instrument will only provide us with information in the complier subpopulation.

- Notice that the effect in this subpopulation doesn't need to be the same as that in the whole sample!

Framework, cont.

- Potential outcomes framework. Potential outcome of individual i, depending on value of the treatment, D:

$$Y_{1i} \quad \text{if} \quad D_i = 1$$

$$Y_{0i} \quad \text{if} \quad D_i = 0$$

- The parameters of interest are defined as follows:

$$q_\tau(Y_i | X_i, D_i, D_{1i} > D_{0i}) = \alpha_\tau D_i + X_i' \beta_\tau, \quad (17)$$

where:

  - $q_\tau(Y_i | X_i, D_i, D_{1i} > D_{0i})$: τ quantile of Y_i given X_i and D_i and conditional on being a complier, $D_{1i} > D_{0i}$

  - α_τ, β_τ: quantile regression coefficients for compliers

Interpretation of α_τ:

- Recall that in the population of compliers ($D_{1i} > D_{0i}$) and conditional on X, D is independent of potential outcomes.

- Why? the instrument Z is a source of exogenous variation in treatment status in this group.

- Therefore, α_τ: Difference in the conditional-on-X quantiles of the treated (Y_{1i}) and non-treated (Y_{0i}) for compliers ($D_{1i} > D_{0i}$).

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

- What α_τ is NOT measuring

1. This is not a comparison between individuals who effectively received the treatment (for instance, attended college), and individuals who did not (i.e., unconditional distribution of Y). The results are conditional on X!

2. We're not estimating the conditional quantile of the individual treatment effects: $q_\tau(Y_{1i} - Y_{0i})$. Unlike in the conditional mean case the difference of quantiles is not the quantile of the difference!

- Let's consider this last point a bit more:

- When estimating conditional expectations: the mean of the differences is the differences of the means

- In quantiles: this is not true, the quantile of the difference is not the difference of the quantiles!

- Therefore: we'll be comparing the (conditional) distribution of treated and the distribution of not treated, we're not comparing individuals.

- As we saw in the previous section, we would need to impose strong conditions so that these functions are the same (monotonicity conditions, which are related to the rank invariance of a treatment).

- But typically knowing the difference of the quantiles is enough.

- Why? consider a training program. For evaluation purposes it would be enough if we observe that the people that took the program are better off.

The QTE Estimator

- Key idea: Z is a source of exogenous variation (i.e., conditional on X it's unrelated to U). Quantile regression coefficients can (theoretically) be estimated by running QR in the population of compliers.

- Problem: We do not observe whether an individual is a complier or not.

- Solution: Let's look for the compliers. To do that, we'll use Abadie (2003) "Kappa" theorem to find them.

Main idea:

- I'd like to estimate the effect of treatment by comparing treated and non treater individuals in the complier population

- For this I need to "find" the non-compliers and remove them from the comparison group.

- The latter individuals are of two types:
  - Always takers
  - Never takers.

- Let's define an operator κ_i that "finds compliers".

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

- Intuition:

  - individuals with $D_i(1-Z_i) = 1$ are always-takers as, for this term to be 1 then $D_i = 1$ and $Z_i = 0$.

  - individuals with $(1-D_i)Z_i = 1$ are never-takers, as $1 - D_i = 1$ (i.e., $D_i = 0$) and $Z_i = 1$;

  - hence, the left-out are the compliers!

  - Indeed, it can be checked that

$$E[\kappa_i \mid Y_i, X_i, D_i] = \Pr(D_{1i} > D_{0i} \mid Y_i, X_i, D_i).$$

- Abadie's (2000) result:

- Let $g(Y_i, X_i, D_i)$ be any measurable function of (Y_i, X_i, D_i) with finite expectation, and Z_i be a binary instrument that satisfies the standard assumptions given X_i, then:

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

where:

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

- Given this result, Abadie, Angrist, and Imbens (2002) developed the QTE estimator as the sample analogue of:

$$(\alpha\tau, \beta\tau') = \arg\min_{a,b} E[\rho_\tau(Y_i - aD_i - X_i'b) \mid D_{1i} > D_{0i}]) \qquad (3)$$

$$= \arg\min_{a,b} E[\kappa_i \rho_\tau(Y_i - aD_i - X_i'b)] \qquad (4)$$

- To obtain the estimator, substitute expectation by sample mean.

Practical considerations

- κ_i needs to be estimated. The uncertainty in the estimation of this parameter has an impact on the distribution of the estimators of the main parameter (α_τ).

- Typically, bootstrap is employed (including the estimation of κ_i in the bootstrapping). Abadie et al. (2002) also provide the asymptotic distribution, but it's less employed.

- To avoid non-convexities in the optimization process, in practice, expresion (4) above is replaced by this one:

$$(\alpha\tau, \beta\tau') = \arg\min_{a,b} E[E(\kappa_i|Y_i, D_i, X_i)\rho_\tau(Y_i - aD_i - X_i'b) \tag{5}$$

(which is obtained by iterating expectations in (4))

- A further simplification gives:

$$E[\kappa_i \mid Y_i, X_i, D_i] =$$

$$1 - \frac{D_i(1 - E[Z_i \mid Y_i, X_i, D_i = 1])}{1 - \Pr(Z_i = 1 \mid X_i)} - \frac{(1 - D_i)E[Z_i \mid Y_i, X_i, D_i = 0]}{1 - \Pr(Z_i = 1 \mid X_i)} \quad (6)$$

- this is the expression used in the QTE estimator

- A very simple to implement the QTE estimator consists of the following two steps:

  1. Estimate $E[\kappa_i \mid Y_i, X_i, D_i]$
  2. Perform quantile regression on $\rho_\tau(Y_i - aD_i - X_i'b)$ (e.g., with `qreg`) using these predicted κ's as weights.

- How to estimate $E[\kappa_i \mid Y_i, X_i, D_i]$:

- See details in Mostly Harmless, p. 287

- It's done by running some probit regressions of a) Z_i on Y_i and X_i for D=1 and D=0, (separately)

b) A probit of Z_i on X_i (whole sample)

- Construct $E(\kappa_i|Y_i, D_i, X_i)$ by replacing (a) and (b) in (6) above.

- Fortunately, we can also do all this using a very recent STATA user-written command

An example

- From Abadie et al, 2003.
- Job Training partnership Act (JTPA): large federal program providing subsidized training to disadvantaged american workers (randomly assigned)
- Effect of the program?
- Sample: 5102 adult mean with 30 month earnings data in the sample.
- Key variables:
- Y_i :earnings
- D_i: training received
- Z_i: randomly assigned offer of training program

- Problem: some participants declined the intervention being offered (only 60% of the potential participants accepted the training)

- Thus: treatment received (D) is not random! it's therefore partly self-selected and likely to be correlated with potential individual characteristics, and then, potential outcomes.

- Instrument: offer received to participating in the program

- Covariates: Since Z is truly random, covariates are not really needed to estimate the effects on compliers. However, even in these type of situations it's customary to control for other variables to correct for chance associations and to increase precision.

- Following TAble: OLS and QR, (first panel), 2SLS and QTE

TABLE 7.2.1

Quantile regression estimates and quantile treatment effects from the JTPA experiment

A. OLS and Quantile Regression Estimates

Variable	OLS	Quantile .15	.25	.50	.75	.85
Training effect	3,754	1,187	2,510	4,420	4,678	4,806
	(536)	(205)	(356)	(651)	(937)	(1,055)
% Impact of training	21.2	135.6	75.2	34.5	17.2	13.4
High school or GED	4,015	339	1,280	3,665	6,045	6,224
	(571)	(186)	(305)	(618)	(1,029)	(1,170)
Black	−2,354	−134	−500	−2,084	−3,576	−3,609
	(626)	(194)	(324)	(684)	(1087)	(1,331)
Hispanic	251	91	278	925	−877	−85
	(883)	(315)	(512)	(1,066)	(1,769)	(2,047)
Married	6,546	587	1,964	7,113	10,073	11,062
	(629)	(222)	(427)	(839)	(1,046)	(1,093)
Worked < 13 weeks in past year	−6,582	−1,090	−3,097	−7,610	−9,834	−9,951
	(566)	(190)	(339)	(665)	(1,000)	(1,099)
Constant	9,811	−216	365	6,110	14,874	21,527
	(1,541)	(468)	(765)	(1,403)	(2,134)	(3,896)

B. 2SLS and QTE Estimates

Variable	2SLS	Quantile .15	.25	.50	.75	.85
Training effect	1,593	121	702	1,544	3,131	3,378
	(895)	(475)	(670)	(1,073)	(1,376)	(1,811)
% Impact of training	8.55	5.19	12.0	9.64	10.7	9.02
High school or GED	4,075	714	1,752	4,024	5,392	5,954
	(573)	(429)	(644)	(940)	(1,441)	(1,783)
Black	−2,349	−171	−377	−2,656	−4,182	−3,523
	(625)	(439)	(626)	(1,136)	(1,587)	(1,867)
Hispanic	335	328	1,476	1,499	379	1,023
	(888)	(757)	(1,128)	(1,390)	(2,294)	(2,427)
Married	6,647	1,564	3,190	7,683	9,509	10,185
	(627)	(596)	(865)	(1,202)	(1,430)	(1,525)
Worked <13 weeks in past year	−6,575	−1,932	−4,195	−7,009	−9,289	−9,078
	(567)	(442)	(664)	(1,040)	(1,420)	(1,596)
Constant	10,641	−134	1,049	7,689	14,901	22,412
	(1,569)	(1,116)	(1,655)	(2,361)	(3,292)	(7,655)

Notes: The table reports OLS, quantile regression, 2SLS, and QTE estimates of the effect of training on earnings (adapted from Abadie, Angrist, and Imbens, 2002). The sample includes 5,102 adult men. Assignment status is used as an instrument for training status in Panel B. In addition to the covariates shown in the table, all models include dummies for service strategy recommended and age group, and a dummy indicating data from a second follow-up survey. Robust standard errors are reported in parentheses.

- Evidence of positive selection (compare the OLS and 2SLS, for instance)

- Very large effects for lower quantiles in QR but very low after instrumenting!

- Effect is concentrated in the upper quantiles!

An example using STATA

Using geographic variation in college proximity to estimate the return to schooling (Card, 1993)

- Goal: impact of college attendance on wages

- Key variable: college attendance (dummy)

- Problem: it's endogeneous, (college attendance is correlated with unobserved variables, for instance, ability, socio-economic status, etc).

- Instrument: college proximity

- Card showed that people (his sample only had men, in fact) who were raised in local labor markets had significantly higher levels of education, even controlling for background factors (parental education, etc).

- We will use the IVQTE package to compute the QTE estimator (but remember, you can also compute it following the steps mentioned above using just probit and qreg!)

- Then, dep variable is log wages, indep. variable is college attendance, controls: mother's education, experience, region and black (dummy)

■ Quantile regression (no instrumenting yet)

■ We can get the same estimates using qreg and ivqte! (no instrumenting yet). Let's see that: qreg lwage college exper black motheduc reg662 reg663 reg664 reg665 reg666

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
college	.0787956	.0393212	2.00	0.045	.0016852	.155906
age	.0405845	.0057545	7.05	0.000	.0292997	.0518692
black	-.2602717	.0534961	-4.87	0.000	-.3651797	-.1553637
fatheduc	-.0079807	.0063746	-1.25	0.211	-.0204816	.0045202
motheduc	.0179566	.007538	2.38	0.017	.0031743	.0327389
reg662	.1183562	.0925071	1.28	0.201	-.063054	.2997663
reg663	.190697	.0912201	2.09	0.037	.0118107	.3695833
reg664	.0225945	.1064365	0.21	0.832	-.1861316	.2313207
reg665	.0177978	.0937132	0.19	0.849	-.1659776	.2015732
reg666	-.049373	.1053505	-0.47	0.639	-.2559694	.1572235
reg667	-.0109539	.0995215	-0.11	0.912	-.2061196	.1842118
reg668	.0037395	.1289615	0.03	0.977	-.2491591	.2566381
reg669	.0644794	.0997591	0.65	0.518	-.1311521	.2601109
_cons	4.488972	.2034505	22.06	0.000	4.089998	4.887947

```
. *the same point estimates but different standard errors (consistent in case of heterosced
> asticity) are obtained with ivqte
```

ivqte lwage exper black motheduc reg662 reg663 reg664 reg665 reg666 reg667 reg668 reg669 (college), quantiles(0.1) variance

```
. ivqte lwage age black fatheduc motheduc reg662 reg663 reg664 reg665 reg666 reg667 reg668
> reg669 (college), q(0.1) variance

Quantile regression
Estimator suggested in Koenker and Bassett (1978)

Quantile:                     .1
Dependent variable:           lwage
Regressor(s):                 college age black fatheduc motheduc reg662 reg663 reg664 reg66
> 5 reg666 reg667 reg668 reg669
Number of observations:       2220
```

lwage	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
college	.0787956	.0374622	2.10	0.035	.005371	.1522202
age	.0405845	.0048544	8.36	0.000	.0310701	.0500989
black	-.269386	.0470991	-5.72	0.000	-.3616986	-.1770733
fatheduc	-.0079807	.0054604	-1.46	0.144	-.0186829	.0027215
motheduc	.0179566	.0066357	2.71	0.007	.0049508	.0309624
reg662	.1183562	.086721	1.36	0.172	-.0516139	.2883262
reg663	.190697	.0860994	2.21	0.027	.0219453	.3594487
reg664	.0225945	.0938826	0.24	0.810	-.161412	.2066011
reg665	.026912	.0836184	0.32	0.748	-.136977	.1908011
reg666	-.0402587	.1028796	-0.39	0.696	-.2418991	.1613816
reg667	-.0109539	.0865353	-0.13	0.899	-.1805599	.1586521
reg668	.0037395	.1178051	0.03	0.975	-.2271543	.2346333
reg669	.0644794	.0992441	0.65	0.516	-.1300354	.2589943
_cons	4.488972	.1560393	28.77	0.000	4.183141	4.794804

.

- Point estimates are exactly identical (because ivqte calls qreg) BUT the standard errors differ

- Standard errors of ivqte are preferred, they are robust against heteroskedasticity and other forms of dependence between the residuals and the regressors.

- Abadie, Angrist and Imbens estimator

```
. ivqte lwage (college=nearc4), q(0.1) variance dummy(black) continuous(age fatheduc mothed
> uc) unordered(region) aai

IV quantile regression
Estimator suggested in Abadie, Angrist and Imbens (2002)

Quantile(s):                 .1
Dependent variable:          lwage
Treatment variable:          college
Instrumental variable:       nearc4
Control variable(s):         age fatheduc motheduc black region
Number of observations:      2220
Proportion of compliers:     .074

Propensity score estimated by local logit regression with h = infinity and lambda = 1
Positive weights estimated by local linear regression with h = infinity and lambda = 1
Variance estimated using local linear regression with h = infinity and lambda = 1
```

lwage	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
college	.636274	.2713248	2.35	0.019	.1044873	1.168061
age	.0670265	.0677152	0.99	0.322	-.0656929	.199746
fatheduc	-.0005916	.0813788	-0.01	0.994	-.1600911	.1589079
motheduc	.00345	.0693573	0.05	0.960	-.1324877	.1393877
black	-.1726069	.6434898	-0.27	0.789	-1.433824	1.08861
region2	.8507937	.4571578	1.86	0.063	-.045219	1.746806
region3	.8496646	.4607969	1.84	0.065	-.0534808	1.75281
region4	.830908	.556047	1.49	0.135	-.2589242	1.92074
region5	.8543029	.762362	1.12	0.262	-.6398993	2.348505
region6	.7592364	1.216877	0.62	0.533	-1.625798	3.144271
region7	.7541343	1.049167	0.72	0.472	-1.302194	2.810463
region8	.4590159	.8000379	0.57	0.566	-1.109029	2.027061
region9	.8812575	.7842898	1.12	0.261	-.6559223	2.418437
_cons	2.67033	2.722238	0.98	0.327	-2.665159	8.005819

Takeaways

- Using QR we can investigate the effects of covariates not only in the central values of the distribution, but also in the tails on in any other point we might be interested in

- All the aspects we studied in conditional mean estimation can be re-studied here: very large literature!

- Still some unresolved issues, literature is still active in this area!