Lecture 13 – Quantile regression, Quantile Treatment effects

Economics 8379
George Washington University

Instructor: Prof. Ben Williams
Introduction

Quantiles and QTEs

Conventional Quantile Regression

Censored QR

QTEs under CI

QTEs with IV
Introduction

- Typically there is a focus on means and conditional means in econometrics.
  - In the program evaluation literature, objects of interest are typically means:
    
    $$ATE = E(Y_1 - Y_0)$$
    $$TT = E(Y_1 - Y_0 \mid D = 1)$$

    in addition to LATE, PRTE, MPRTE, etc.
Introduction

- Typically there is a focus on means and conditional means in econometrics.
  - In the program evaluation literature, objects of interest are typically means:
    
    \[ \text{ATE} = E(Y_1 - Y_0) \]
    \[ TT = E(Y_1 - Y_0 \mid D = 1) \]

    in addition to LATE, PRTE, MPRTE, etc.
  - In regression model, \( Y_i = \beta' X_i + \varepsilon_i \), identification and estimation are based on conditional mean condition that \( E(\varepsilon_i \mid X_i) = 0 \).
Introduction

• Typically there is a focus on means and conditional means in econometrics.
  • In the program evaluation literature, objects of interest are typically means:
    \[
    ATE = E(Y_1 - Y_0) \\
    TT = E(Y_1 - Y_0 \mid D = 1)
    \]
    in addition to LATE, PRTE, MPRTE, etc.
  • In regression model, \( Y_i = \beta' X_i + \varepsilon_i \), identification and estimation are based on conditional mean condition that \( E(\varepsilon_i \mid X_i) = 0 \).
  • Angrist and Pischke’s discussion of OLS is based on the idea that it provides the linear approximation with the minimum \textit{mean} squared error.
Introduction

- But the distribution of outcomes may shift in response to policy/treatment/price changes, etc.
- This lecture focuses on two aspects.
  - Defining objects of interest that go beyond the mean.
  - Identification and estimation, mostly in non-structural models.
Introduction

Quantiles and QTEs

Conventional Quantile Regression

Censored QR

QTEs under CI

QTEs with IV
Quantile function

- Let $F_Y(y)$ denote the cdf of the random variable $Y$.
  - If $F_Y$ is continuous and strictly increasing then we can define
    \[ Q_\tau(Y) := F_Y^{-1}(\tau) \]
    It follows that $F_Y(Q_\tau(Y)) = \tau$
  - More generally,
    \[ Q_\tau(Y) := \inf\{y : F_Y(y) \geq \tau\} \]
Conditional quantile function

- Let $F_{Y|X}(y \mid x)$ denote the conditional cdf of the random variable $Y$ given $X = x$.
- If $F_{Y|X}(y \mid x)$ is continuous and strictly increasing in $y$ then we can define

$$Q_{\tau}(Y \mid X = x) := F_{Y|X}^{-1}(\tau \mid x)$$

It follows that $F_{Y|X}(Q_{\tau}(Y \mid X = x) \mid x) = \tau$
CQF in some models

- Example 1. \( Y = m(X) + \varepsilon. \)
  - First, \( Q_\tau(Y \mid X) = m(X) + Q_\tau(\varepsilon \mid X). \)
  - Proof:

\[
\Pr(Y \leq m(X) + Q_\tau(\varepsilon \mid X) \mid X = x) \\
= \Pr(\varepsilon \leq Q_\tau(\varepsilon \mid X) \mid X = x) \\
= \tau
\]
CQF in some models

Example 1. $Y = m(X) + \varepsilon$.

- “homoskedasticity”: $Q_\tau(Y \mid X) = m(X) + Q_\tau(\varepsilon)$ so that the CQF is separable in $x$ and $\tau$.

- linear, heteroskedastic:
  - Suppose $m(x) = \beta'x$ and $\varepsilon = (\lambda'X)\tilde{\varepsilon}$ where $\tilde{\varepsilon}$ is independent of $X$.
  - Then
    \[
    Q_\tau(\varepsilon \mid X) = (\lambda'X)Q_\tau(\tilde{\varepsilon})
    \]
    so that
    \[
    Q_\tau(Y \mid X) = X'(\beta + \lambda Q_\tau(\tilde{\varepsilon}))
    \]
CQF in some models

• Example 2. Suppose $X$ is a binary “treatment” and

$$Y = \beta_0 + (\beta_1 - \beta_0 + U_1 - U_0)X + U_0$$

where $U_0 \perp\!\!\!\!\perp X$ and $U_1 \perp\!\!\!\!\perp X$
Example 2. Suppose $X$ is a binary “treatment” and

$$Y = \beta_0 + (\beta_1 - \beta_0 + U_1 - U_0)X + U_0$$

where $U_0 \perp\!\!\!\!\perp X$ and $U_1 \perp\!\!\!\!\perp X$

Then

$$Q_\tau(Y \mid X) = \beta_0 + Q_\tau(U_0) + X(\beta_1 - \beta_0 + Q_\tau(U_1) - Q_\tau(U_0))$$
Example 3. Suppose $Y = g(X, U)$ where $U \mid X \sim \text{Uniform}(0, 1)$ and $g(x, u)$ is strictly monotonic in $u$.

This is a general nonseparable model for $Y$.

In this model,

$$\Pr(Y \leq g(x, \tau) \mid X = x) = \Pr(g(x, U) \leq g(x, \tau) \mid X = x) = \tau$$

So

$$Q_{\tau}(Y \mid X) = g(X, \tau)$$
linear CQF model

- Rather than starting with an equation for $Y$, the linear quantile regression model assumes that

$$ Q_\tau(Y \mid X) = \beta(\tau)'X $$

- The linear, heteroskedastic model is one model that produces this form for the CQF.
CQF to QF

- What if we care about the unconditional quantile function?
- The law of iterated expectations now looks really nice!

\[ E(Y) = E(E(Y \mid X)) \]

and

\[ E(Y \mid X_1) = E(E(Y \mid X_1, X_2) \mid X_1) \]
CQF to QF

- What if we care about the unconditional quantile function?
- The law of iterated expectations now looks really nice!

\[ E(Y) = E(E(Y \mid X)) \]

and

\[ E(Y \mid X_1) = E(E(Y \mid X_1, X_2) \mid X_1) \]

- But \( Q_\tau(Y) \neq E(Q_\tau(Y \mid X)) \) or any other simple functional of \( Q_\tau(Y \mid X = x) \).
- Instead,

\[ F_Y(y) = E \left( \int_0^1 1(Q_\tau(Y \mid X) < y) d\tau \right) \]

and \( Q_\tau(Y) = F_Y^{-1}(\tau) \).
Treatment effects

- Let $\Delta = Y_1 - Y_0$ denote the individual level treatment effect. Then $F_\Delta(\delta)$ and $Q_\tau(\Delta)$ represent the distribution of treatment effects.
- The quantile treatment effect is different:

$$QTE(\tau) = Q_\tau(Y_1) - Q_\tau(Y_0)$$

- The QTE represents the effect of treatment on the $\tau^{th}$ quantile of the outcome distribution.
- Note that it's possible that $QTE(\tau) = 0$ for all $\tau$ but $Pr(\Delta = 0) = 0$. 
Treatment effects

- Example of QTE estimate from Andrews, Li, Lovenheim (2012):

Figure 7: Quantile Treatment Effects of Graduating from UT-Austin on Earnings by Race

Source: Authors' calculations from the University of Texas at Dallas Education Research Center data and administrative earnings records as described in the text. Each estimated point is the difference between the observed earnings at each percentile for UT-Austin (Panel A), Texas A&M (Panel B) and community colleges (Panel C) and the associated earnings at that percentile from the counterfactual earnings distribution. The dotted lines show the bounds of the 95% confidence intervals for each percentile point. The horizontal dashed lines show the 95% confidence interval of the mean effect from Table 4.
Treatment effects

- Example of QTE estimate from Andrews, Li, Lovenheim (2012):

Figure 9: Quantile Treatment Effects of Graduating from a Community College on Earnings by Race

Panel A: White

Panel B: Black

Panel C: Asian

Panel D: Hispanic

Source: Authors' calculations from the University of Texas at Dallas Education Research Center data and administrative earnings records as described in the text. Each estimated point is the difference between the observed earnings at each percentile for UT-Austin (Panel A), Texas A&M (Panel B) and community colleges (Panel C) and the associated earnings at that percentile from the counterfactual earnings distribution. The dotted lines show the bounds of the 95% confidence intervals for each percentile point. The horizontal dashed lines show the 95% confidence interval of the mean effect from Table 4.
Treatment effects

- Estimating $Q_{\tau}(\Delta)$ requires assumptions allowing us to identify the dependence between $Y_0$ and $Y_1$.
- Estimating the QTE is much simpler as it requires only knowledge of the marginal distributions of $Y_0$ and $Y_1$. 
Censoring and other tail problems

- Suppose we know $F_Y(y)$ for $y \leq \bar{y}$ but not for $y \geq \bar{y}$.
  - What can be said about $E(Y)$?
  - When can we determine $Med(Y)$?
  - More generally, can we determine $Q_\tau(Y)$?
Censoring and other tail problems

- Suppose we know $F_Y(y)$ for $y \leq \bar{y}$ but not for $y \geq \bar{y}$.
  - What can be said about $E(Y)$?
  - When can we determine $Med(Y)$?
  - More generally, can we determine $Q_\tau(Y)$?
    - Yes! For $\tau \leq F_Y(\bar{y})$. 

Suppose we know $F_Y(y)$ for $y \leq \bar{y}$ but not for $y \geq \bar{y}$.

- What can be said about $E(Y)$?
- When can we determine $Med(Y)$?
- More generally, can we determine $Q_\tau(Y)$?
  - Yes! For $\tau \leq F_Y(\bar{y})$.

What if we observe $Y$ only if $\underline{y} \leq Y \leq \bar{y}$?

- Then $Q_\tau(Y)$ is identified for all $F_Y(\underline{y}) \leq \tau \leq F_Y(\bar{y})$. 

Censoring and other tail problems
Introduction

Quantiles and QTEs

**Conventional Quantile Regression**

Censored QR

QTEs under CI

QTEs with IV
Estimation

- Suppose we have an i.i.d. sample $(Y_i, X_i)$.
- As mentioned above, the conventional model assumes that $Q_\tau(Y_i | X_i) = \beta(\tau)'X_i$.
- This can be estimated through a method of moments because

$$Q_\tau(Y | X = x) = \arg\min_{q_\tau(x)} E(\rho_\tau(Y - q_\tau(x)) | X = x)$$

where $\rho_\tau(u) = 1(u > 0)\tau|u| + 1(u \leq 0)(1 - \tau)|u|$.
- $\rho_\tau$ is called the “check function” and can also be written as $u(\tau - 1(u \leq 0))$.
- In the linear case this implies that

$$\beta_\tau = \arg\min_{b_\tau} E(\rho_\tau(Y - b_\tau'X))$$
Estimation

• The standard quantile regression estimator solves the sample analog of this problem:

\[ \hat{\beta}_\tau = \arg \min_{b_\tau} n^{-1} \sum_{i=1}^{n} \rho_\tau(Y_i - b'_\tau X_i) \]

• This estimator is consistent and asymptotically normal.
Standard errors

- The asymptotic variance of $\hat{\beta}(\tau)$ is $A^{-1}BA^{-1}$ where

$$A = E(f_{u_\tau}(0 \mid X_i)X_iX_i')$$
$$B = \tau(1 - \tau)E(X_iX_i')$$

where $u_\tau = Y - \beta(\tau)'X$.

- Only valid if the CQF is correctly specified.
- Under independence between $u_\tau$ and $X_i$, this simplifies.
Standard errors

- The analytical formula is difficult – requires nonparametric estimation, specification of bandwidth parameter,...
- It is common to use the bootstrap instead.
  - Bootstrap has been shown to lead to consistent SE estimates.
  - No asymptotic refinement.
Interpretation

- The estimates represent the effect of ceteris paribus changes in the associated regressors on the $\tau^{th}$ quantile of the conditional distribution of $Y_i$.
- It does not represent the effect for an individual at the $\tau^{th}$ quantile.
- Note that $H(Q_\tau(Y \mid X)) = Q_\tau(H(Y) \mid X)$ for monotonic transformation $H$. 
Example

- Table 7.1.1 reports results of a quantile regression of wages on education, controlling for race and a quadratic in experience.

Table 7.1.1
Quantile regression coefficients for schooling in the 1980, 1990, and 2000 censuses

<table>
<thead>
<tr>
<th>Census</th>
<th>Obs.</th>
<th>Mean</th>
<th>SD</th>
<th>Quantile Regression Estimates</th>
<th>OLS Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
<td>0.25</td>
</tr>
<tr>
<td>1980</td>
<td>65,023</td>
<td>6.4</td>
<td>.67</td>
<td>.074 (0.002)</td>
<td>.074 (0.001)</td>
</tr>
<tr>
<td>1990</td>
<td>86,785</td>
<td>6.5</td>
<td>.69</td>
<td>.112 (0.003)</td>
<td>.110 (0.001)</td>
</tr>
<tr>
<td>2000</td>
<td>97,397</td>
<td>6.5</td>
<td>.75</td>
<td>.092 (0.002)</td>
<td>.105 (0.001)</td>
</tr>
</tbody>
</table>

Notes: Adapted from Angrist, Chernozhukov, and Fernandez-Val (2006). The table reports quantile regression estimates of the returns to schooling in a model for log wages, with OLS estimates shown at the right for comparison. The sample includes U.S.-born white and black men aged 40–49. The sample size and the mean and standard deviation of log wages in each census extract are shown at the left. Standard errors are reported in parentheses. All models control for race and potential experience. Sampling weights were used for the 2000 census estimates.
Example

- Table 7.1.1 reports results of a quantile regression of wages on education, controlling for race and a quadratic in experience.
- Results:
  - From 1980 to 1990, the return to schooling increased from roughly 7% to 11%, across all quantiles.
  - From 1990 to 2000, the effect of education on the conditional median of the earnings distribution was unchanged.
  - In 2000, $\hat{\beta}(\tau)$ varies from 9% to 16%.
Example

- Replicating Table 7.1.1.
- Results:
  - The command `qreg` is fairly straightforward:
    \[
    \text{qreg logwk educ black exper exper2, q(90) vce(robust)}
    \]
  - `bsqreg` computes bootstrap standard errors:
    \[
    \text{bsqreg logwk educ black exper exper2, q(90) reps(100)}
    \]
  - `sqreg` estimates multiple \( \tau \)'s simultaneously, facilitating testing:
    \[
    \text{sqreg logwk educ black exper exper2, q(50 90) test [q50]educ=[q90]educ}
    \]
Example

- More on QR in Stata:
  - `qreg` allows weights but `sqreg` and `bsqreg` do not:
    ```stata
    qreg logwk educ black exper exper2 [pweight=perwt], q(90)
    ```
  - `vce(robust)` option computes robust (to heteroskedasticity) standard errors
  - `qreg2` computes standard errors that are robust to misspecification
  - `grqreg` is useful for plotting results
Introduction

Quantiles and QTEs

Conventional Quantile Regression

Censored QR

QTEs under CI

QTEs with IV
Random censoring

- Suppose $\tilde{Y}_i = \min\{Y_i, C_i\}$ and only $(\tilde{Y}_i, X_i, C_i)$ is observed.
  - Top-coding example is a special case of this.
- In addition, assume that $Q_\tau(Y_i \mid X_i, C_i) = \beta(\tau)'X_i$.
- Then $Q_\tau(\tilde{Y}_i \mid X_i) = \min\{\beta(\tau)'X_i, C_i\}$.
- The Powell (1986) censored quantile regression (CQR) estimator solves

$$
\hat{\beta}_\tau = \arg\min_{b_\tau} n^{-1} \sum_{i=1}^n \rho_\tau(Y_i - \min\{\beta(\tau)'X_i, C_i\})
$$
Random censoring

- This estimator performs poorly in many samples with small $n$ and lots of censoring.
- The CQR estimator throws away observations with $Q_\tau(Y_i \mid X_i, C_i) > C_i$.
- Note that this is equivalent to the condition, $1 - \tau < Pr(Y_i > C_i \mid X_i, C_i)$.
  - If we have a correctly specified model for $Pr(Y_i > C_i \mid X_i, C_i)$. (Buchinsky and Hahn (1998) and Khan and Powell (2001).
  - If we don’t. (Chernozhukov and Hong (2002))
Random censoring

- Additional difficulties if:
  - $C_i$ is only observed when $C_i \leq Y_i$
  - $C_i$ is endogenous
- Note that we have discussed parametric solutions to these problems already. (Tobit, Heckit, Roy model, etc.)
- See Portnoy (2003), Koenker (2005) for the first problem.
- See Khan, Ponomareva, and Tamer (2013) for a solution to the second problem in a panel data context.
Introduction

Quantiles and QTEs

Conventional Quantile Regression

Censored QR

QTEs under CI

QTEs with IV
Identification of the ATE

• Recall how the ATE is identified when $Y_1$, $Y_0$ are independent of $D$ conditional on $X$.
  • First, $E(Y_d | X) = E(Y | D = d, X)$.
  • By the law of iterated expectations though, $E(Y_d) = E(E(Y_d | X))$.
  • Therefore,

$$ATE = E(E(Y | D = 1, X) - E(Y | D = 0, X))$$

• Intuitively,
  • Compare treatment and controls that are “matched” on their $X$ values.
  • Then average over the population of interest.
Identification of the ATE

- Identification of the ATE can alternatively be viewed through a reweighting approach.
  - Note that
    \[ E(Y_1) = E \left( \frac{D}{P(X)} Y \right) \]
    where \( P(X) := Pr(D = 1 \mid X) \).
  - Similarly,
    \[ E(Y_0) = E \left( \frac{1 - D}{1 - P(X)} Y \right) \]
    where \( P(X) := Pr(D = 1 \mid X) \).
Identification of the QTE

- We could first try an analogy of the first ATE argument.
  - $Y_1$ and $Y_0$ are independent of $D$ conditional on $X$.
  - So $F_{Y|D=1,X}(y|x) = F_{Y_1|X}(y|x)$.
  - So $F_{Y_1}(y) = \int F_{Y_1|X}(y|x)f_X(x)dx$.
  - After averaging, this can be inverted to get the quantile.
  - Same steps repeated for $D = 0$. 
Firpo (2007) suggests the following alternative approach.

Consider

$$E \left( \frac{D}{P(X)} \mathbf{1}(Y \leq y) \right)$$

This is equal to $F_{Y_1}(y)$.

So solve $\tau = E \left( \frac{D}{P(X)} \mathbf{1}(Y \leq q_{1,\tau}) \right)$.

Equivalently, solve

$$q_{1,\tau} = \arg \min_q E \left( \frac{D}{P(X)} \rho_\tau(Y - q) \right)$$
Identification of the QTE

- Estimation follows in four steps.
  - First estimate the propensity score, $\hat{P}(X_i)$.
  - Second, estimate
    \[
    \hat{q}_{1,\tau} = \arg\min_q n^{-1} \sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)} \rho_\tau(Y_i - q)
    \]
  - Third, estimate
    \[
    \hat{q}_{0,\tau} = \arg\min_q n^{-1} \sum_{i=1}^{n} \frac{1 - D_i}{1 - \hat{P}(X_i)} \rho_\tau(Y_i - q)
    \]
  - Finally, $\hat{QTE}_\tau = \hat{q}_{1,\tau} - \hat{q}_{0,\tau}$.
Identification of the QTE

- This can then be implemented manually in Stata using `qreg with pweights`.
- There is also a new command `poparms` that can implement this in Stata.
Interpretation

- Two approaches:
  - Abadie, Angrist, and Imbens (2002)
    - uses LATE assumptions to identify the QTE for *compliers*
    - *ivqte* in Stata
  - Chernozhukov and Hansen (2007)
    - identify the QTE for the whole population under stronger assumptions
    - *ivqreg* in Stata