Lecture 5: Bad Controls, Standard Errors, Quantile Regression

Fabian Waldinger

Topics Covered in Lecture

1. Bad controls.
2. Standard errors.
3. Quantile regression.

Bad Control Problem

- Controlling for additional covariates increases the likelihood that regression estimates have a causal interpretation.
- More controls are not always better: "bad control problem".
- Bad controls are variables that could themselves be outcomes.
- The bad control problem is a version of selection bias.
- The problem is common in (bad) empirical work.

An Example

- We are interested in the effect of a college degree on earnings.
- People can work in two occupations: white collar ($w_i = 1$) or blue collar ($w_i = 0$).
- A college degree also increases the chance of getting a white collar job.
  $\Rightarrow$ Potential outcomes of getting a college degree: earnings and working in a white collar job.
- Suppose you are interested in the effect of a college degree on earnings. It may be tempting to include occupation as an additional control (because earnings may substantially differ across occupations):

$$Y_i = \beta_1 + \beta_2 C_i + \beta_3 W_i + \varepsilon_i$$

- Even if getting a college degree is randomly assigned one would not estimate the causal effect of college on earnings if one controls for occupation.

- Obtaining a college degree C affects both earnings and occupation:

$$Y_i = C_i Y_{1i} + (1 - C_i) Y_{0i}$$
$$W_i = C_i W_{1i} + (1 - C_i) W_{0i}$$

- We assume that C_i is randomly assigned. We can therefore estimate the causal effect of C_i on either Y_i or W_i because independence insures:

$$E[Y_i | C_i = 1] - E[Y_i | C_i = 0] = E[Y_{1i} - Y_{0i}]$$
$$E[W_i | C_i = 1] - E[W_i | C_i = 0] = E[W_{1i} - W_{0i}]$$

- We can estimate these average treatment effects by regressing Y_i and W_i on C_i.

- Consider the difference in mean earnings between college graduates and others conditional on working in a white collar job.
- This can be estimated by regressing Y_i on C_i in a sample where $W_i = 1$:

$$E[Y_i|W_i = 1, C_i = 1] - E[Y_i|W_i = 1, C_i = 0]$$
$$= E[Y_{1i}|W_{1i} = 1, C_i = 1] - E[Y_{0i}|W_{0i} = 1, C_i = 0]$$

- By the joint independence of $\{Y_{1i}, W_{1i}, Y_{0i} W_{0i}\}$ and C_i we get:

$$= E[Y_{1i}|W_{1i} = 1] - E[Y_{0i}|W_{0i} = 1]$$

i.e. expected earnings for people with a college degree in a white collar occupation minus the expected earnings for people without a college degree in a white collar occupation.

- We therefore have something similar to a selection bias:

$$E[Y_{1i}|W_{1i}=1]-E[Y_{0i}|W_{0i}=1]$$

$$=\underbrace{E[Y_{1i}-Y_{0i}|W_{1i}=1]}_{Causal\ Effect}+\underbrace{E[Y_{0i}|W_{1i}=1]-E[Y_{0i}|W_{0i}=1]}_{Selection\ Bias}$$

- In words: The difference in wages between those with and without college conditional on working in a white collar job equals:
  - the causal effect of college on those with $W_{1i}=1$ (people who work in a white collar job if they have a college degree) and
  - selection bias that reflects the fact that college changes the composition of white collar workers.

Standard Errors

Standard Errors in Practical Applications

- From your econometrics course you remember that the variance-covariance matrix of $\widehat{\beta}$ is given by;

$$V(\widehat{\beta}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

- With the assumption of no heteroscedasticity and no autocorrelation $(\Omega = \sigma^2 I)$ this simplifies to:

$$V(\widehat{\beta}) = \sigma^2(X'X)^{-1}$$

- In a lot of applied work, however, the assumption $\Omega = \sigma^2 I$ is not credible.
- We often estimate models that use regressors that vary at a more aggregate level than the data.
- One example: we estimate how test scores of individual students are affected by variables that vary at the class or school level.

Grouped Error Structure

- E.g. In Krueger's (1999) class size paper he has data on individual level test scores but class size only varies at the class level.
- Even though students were assigned randomly to classes in the STAR experiment, the STAR data are unlikely to be independent across observations within classes because:
  - students in the same class share the same teacher.
  - students share the same shocks to learning.
- A more concrete example:

$$Y_{ig} = \beta_0 + \beta_1 x_g + \varepsilon_{ig}$$

- i: individual

- g: group, with G groups.

Grouped Error Structure

- We assume that the residual has a group structure:

$$e_{ig} = v_g + \eta_{ig}$$

- An error structure like this can increase standard errors (Kloek, 1981, Moulton, 1986).
- With this error structure the intraclass correlation coefficient becomes:

$$\rho_e = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}$$

- How would an error structure like that affect standard errors?

Derivation of the Moulton Factor - Notation

$$y_g = \begin{bmatrix} Y_{1g} \\ Y_{2g} \\ \ldots \\ Y_{n_g g} \end{bmatrix} \qquad e_g = \begin{bmatrix} e_{1g} \\ e_{2g} \\ \ldots \\ e_{n_g g} \end{bmatrix}$$

$\swarrow$ stack all y_g's to get y

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_G \end{bmatrix} \qquad x = \begin{bmatrix} \iota_1 x_1 \\ \iota_2 x_2 \\ \ldots \\ \iota_G x_G \end{bmatrix} \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \ldots \\ e_G \end{bmatrix}$$

where ι_1 is a column vector of n_g ones and G is the number of groups

Derivation of the Moulton Factor - Group Covariance

$$E(ee') = \Psi = \begin{bmatrix} \Psi_1 & 0 & \ldots & 0 \\ 0 & \Psi_2 & & .. \\ .. & & .. & 0 \\ 0 & \ldots & 0 & \Psi_G \end{bmatrix}_{(Gxn_g)x(Gxn_g)}$$

$$\Psi_g = \sigma_e^2 \begin{bmatrix} 1 & \rho_e & \ldots & \rho_e \\ \rho_e & 1 & & .. \\ .. & & .. & \rho_e \\ \rho_e & \ldots & \rho_e & 1 \end{bmatrix}_{n_g x n_g}$$

- where $\rho_e = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}$ (see above).
- The dimensions of the matrix are only correct if n_g is the same across all groups.

Derivation of the Moulton Factor

- Rewriting:

$$X'X = \sum_g n_g x_g x_g' \quad \text{and} \quad X'\Psi X = \sum_g x_g \iota_g' \Psi_g \iota_g x_g'$$

- We can calculate that:

$$\Psi_g \iota_g = \sigma_e^2 \begin{bmatrix} 1 & \rho_e & \ldots & \rho_e \\ \rho_e & 1 & & .. \\ .. & & .. & \rho_e \\ \rho_e & \ldots & \rho_e & 1 \end{bmatrix}_{n_g x n_g} \text{x} \begin{bmatrix} 1 \\ 1 \\ \ldots \\ 1 \end{bmatrix}_{n_g x 1}$$

$$= \sigma_e^2 \begin{bmatrix} 1+\rho_e+\rho_e+\ldots+\rho_e \\ 1+\rho_e+\rho_e+\ldots+\rho_e \\ .. \\ 1+\rho_e+\rho_e+\ldots+\rho_e \end{bmatrix}_{n_g x 1} = \sigma_e^2 \begin{bmatrix} 1+(n_g-1)\rho_e \\ 1+(n_g-1)\rho_e \\ .. \\ 1+(n_g-1)\rho_e \end{bmatrix}_{n_g x 1}$$

Derivation of the Moulton Factor

- Using this last result we get:

$$x_g \iota_g' \Psi_g \iota_g x_g' = \sigma_e^2 x_g \iota_g' \begin{bmatrix} 1+(n_g-1)\rho_e \\ 1+(n_g-1)\rho_e \\ .. \\ 1+(n_g-1)\rho_e \end{bmatrix} x_g'$$

- Using $\iota_g' \begin{bmatrix} 1+(n_g-1)\rho_e \\ 1+(n_g-1)\rho_e \\ .. \\ 1+(n_g-1)\rho_e \end{bmatrix} = n_g[1+(n_g-1)\rho_e]$ we get:

$$x_g \iota_g' \Psi_g \iota_g x_g' = \sigma_e^2 n_g [1+(n_g-1)\rho_e] x_g x_g'$$

Derivation of the Moulton Factor

- Now define $\tau_g = 1 + (n_g - 1)\rho_e$ so we get:

$$x_g \iota_g' \Psi_g \iota_g x_g' = \sigma_e^2 n_g \tau_g x_g x_g'$$

- And therefore:

$$X'\Psi X = \sigma_e^2 \sum_g n_g \tau_g x_g x_g'$$

- We can therefore rewrite the variance-covariance matrix with grouped data as:

$$V(\widehat{\beta}) = (X'X)^{-1} X'\Psi X (X'X)^{-1}$$

$$= \sigma_e^2 (\sum_g n_g x_g x_g')^{-1} \sum_g n_g \tau_g x_g x_g' (\sum_g n_g x_g x_g')^{-1}$$

Comparing Group Data Variance to Normal OLS Variance

- If group sizes are equal $n_g = n$ and $\tau_g = \tau$ we get

$$V(\widehat{\beta}) = \sigma_e^2 (\sum_g n x_g x_g')^{-1} \sum_g n\tau x_g x_g' (\sum_g n x_g x_g')^{-1}$$

$$= \sigma_e^2 \tau (\sum_g n x_g x_g')^{-1}$$

- The normal OLS variance-covariance matrix is:

$$V^{ols}(\widehat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma_e^2 (\sum_g n x_g x_g')^{-1}$$

- Therefore $\frac{V^{clustered}(\widehat{\beta})}{V^{OLS}(\widehat{\beta})} = 1 + (n-1)\rho_e$
- This ratio tells us how much we overestimate precision by ignoring the intraclass correlation. The square root of this ratio is called the Moulton factor (Moulton, 1986).

Generalized Moulton Factor

- With equal group sizes the ratio of covariances is (last slide):

$$\frac{V^{clustered}(\widehat{\beta})}{V^{OLS}(\widehat{\beta})} = 1 + (n-1)\rho_e$$

- For a fixed number of total observations this gets larger if group size n goes up and if the within group correlation ρ_e increases.
- The formula above covers the important special case where the group sizes are the same and where the regressors do not vary within groups.
- A more general formula allows the regressor x_{ig} to vary within groups (therefore it is now indexed by i) and it allows for different group sizes n_g :

$$\frac{V^{clustered}(\widehat{\beta})}{V^{OLS}(\widehat{\beta})} = 1 + (\frac{V(n_g)}{\overline{n}} + \overline{n} - 1)\rho_x \rho_e$$

- where $\overline{n}$ is the average group size.

- ρ_x is the intraclass correlation of x_{ig}

Generalized Moulton Factor

- The generalized Moulton formula tells us that clustering has a bigger impact on standard errors when:
  - group sizes vary
  - ρ_x is large, i.e. x_{ig} does not vary much within groups

Clustering and IV

- The Moulton factor works similarly with 2SLS estimates. In that case the Moulton factor is:

$$\frac{V^{clustered}(\widehat{\beta})}{V^{OLS}(\widehat{\beta})} = 1 + (\frac{V(n_g)}{\overline{n}} + \overline{n} - 1)\rho_{\widehat{x}}\rho_e$$

- where $\rho_{\widehat{x}}$ is the intraclass correlation of the first-stage fitted values
- ρ_e is the intraclass correlation of the second-stage residuals

Clustering and Differences-in-Differences

- As discussed in our differences-in-differences lecture we often encounter treatments that take place at the group level.
- In this type of data we have to worry not only about correlation of errors within groups at a particular point in time but also at correlation over time (Bertrand, Duflo, and Mullainathan, 2003).
- In that case errors should be clustered at the group level (not only at the group x time level). For other solutions see differences-in-differences lecture.

Practical Tips for Estimating Models with Clustered Data

1. *Parametric Fix of Standard Errors:*
   Use the generalized Moulton factor formula, calculate $V(n_g)$, ρ_x and ρ_e and adjust standard errors.

2. *Cluster standard errors:*

   The clustered variance-covariance matrix is:

$$V(\widehat{\beta}) = (X'X)^{-1}(\sum_g X_g \widehat{\Psi}_g X_g)(X'X)^{-1}$$

- Where $\widehat{\Psi}_g$ allows for arbitrary correlation of standard errors within groups and is estimated using the residuals.
- This allows not only for autocorrelation but also for heteroscedasticity (more general then the `robust` command)

- $\widehat{\Psi}_g$ is estimated as:

$$\widehat{\Psi}_g = a\widehat{e}_g\widehat{e}_g' = a\begin{bmatrix} \widehat{e}_{1g}^2 & \widehat{e}_{1g}\widehat{e}_{2g} & \cdots & \widehat{e}_{1g}\widehat{e}_{n_g g} \\ \widehat{e}_{1g}\widehat{e}_{2g} & \widehat{e}_{2g}^2 & & \\ \vdots & & \cdots & \widehat{e}_{n_g-1,g}\widehat{e}_{n_g g} \\ \widehat{e}_{1g}\widehat{e}_{n_g g} & \cdots & \widehat{e}_{n_g-1,g}\widehat{e}_{n_g g} & \widehat{e}_{n_g g}^2 \end{bmatrix}$$

- a is a degrees-of-freedom correction.
- The clustered estimator is consistent as the number of groups gets large given any within-group correlation structure and not just the parametric model that we have investigated above. -> Asymptotics work at the group level.
- Increasing group sizes to infinity does not make the estimator consistent if the number of groups does not increase.

3. *Use group averages instead of micro data:*
   I.e. estimate:

$$\overline{Y}_g = \beta_0 + \beta_1 X_g + \overline{e}_g$$

- by WLS using the group size as weights.
- This is equivalent to OLS using the micro data but the standard errors reflect the group structure.
- How do you estimate a model using group averages with regressors that vary at the micro level?

$$Y_{ig} = \beta_0 + \beta_1 x_g + \beta_2 w_{ig} + \varepsilon_{ig}$$

- Estimate: $Y_{ig} = \sum_{gr} \widetilde{Y}_{gr} I(g = gr) + \gamma_2 w_{ig} + \varepsilon_{ig}$

  -> get $\widetilde{Y}_{gr}$
- Estimate: $\widetilde{Y}_{gr} = \beta_0 + \beta_1 x_g + \varepsilon_{ig}$

Practical Tips for Estimating Models with Clustered Data

4. Block bootstrap:
   Draw blocks of data defined by the groups g.
5. Do GLS:
   For this one would have to assume a particular error structure and then estimate GLS.

How Many Clusters Do We Need?

- As discussed above asymtotics work at the group level.
- So what do we do if the cluster count is low?
- The best solution is of course to get data for more groups.
- Sometimes this is impossible, however. What do you do in that case?

Solutions For Small Number of Clusters

1. *Bias correction of clustered standard errors:*
   Bell and McCaffrey (2002) suggest to adjust residuals by:

$$\widehat{\Psi}_g = a\widetilde{e}_g\widetilde{e}_g'$$
$$\widetilde{e}_g = A_g\widehat{e}_g$$

- where A_g solves:
  - $A_g' A_g = (I - H_g)^{-1}$
  - $H_g = X_g(X'X)^{-1}X_g'$
  - and a is a degrees-of-freedom correction

Solutions For Small Number of Clusters

2. *Use t-distribution for inference:*
   Even with the bias adjustment discussed in 1. it is more prudent to use the t-distribution with $G - k$ degrees of freedom rather then the standard normal.
3. *Use estimation at the group mean:*
   Donald and Lang (2007) argue that estimation using group means works well with small G in the Moulton problem. Group estimation calls for a fixed X_g within groups so the group estimation is not a solution in the DiD case (as for that our main regressor of interest varies within groups over time).
4. *Block-boostrap:*
   Cameron, Gelbach and Miller (2008) report that a particular form of a block bootstrap works well with small number of groups.

Quantile Regression

Quantile Regression

- Most of applied work is concerned with averages.
- Applied economists are becoming more interested in what is happening to an entire distribution.
- Suppose you were interested in changes in the wage distribution over time. Mean wages could have remained constant even if wages in the upper quantiles increased while they fell in lower quantiles.
- Quantile regression allows us to analyze these questions.

Quantile Regression

- Suppose we are interested in the distibution of a continuously distributed random variable Y_i with a well-behaved density (no gaps or spikes).
- The conditional quantile function (CQF) at quantile τ given a vector of regressors X_i can be defined as:

$$Q_\tau(Y_i|X_i) = F_y^{-1}(\tau|X_i)$$

where $F_y(y|X_i)$ is the distribution function for Y_i at y, conditional on X_i.

- When $\tau = 0.1$, for example, $Q_\tau(Y_i|X_i)$ describes the lower decile of Y_i given X_i
- When $\tau = 0.5$, for example, $Q_\tau(Y_i|X_i)$ describes the median of Y_i given X_i

The Minimization

- In mean regression (OLS) we minimize the mean-squared error:

$$E[Y_i|X_i] = \underset{m(X_i)}{\arg\min} E[(Y_i - m(X_i))^2]$$

- In quantile regression we minimize:

$$Q_\tau[Y_i|X_i] = \underset{q(X)}{\arg\min} E[\rho_\tau(Y_i - q(X_i))]$$

- With the loss function: $\rho_\tau(u) = (u\ (\tau - I(u < 0))$
- The loss function weights positive and negative terms asymmetrically (unless we evaluate at the median).
- This asymmetric weighting generates a minimand that picks out conditional quantiles (this is not particularly obvious, see Koenker, 2005).

The Loss Function: Example

- Suppose Y_i is a discrete random variable that takes values 1,2,...,9 with equal probabilities.
- We would like to find the median. The expected loss is:

$$E[\rho_\tau(u)] = E[(y-u)\ (\tau - I(y-u<0))]$$

- If $y - u \geqq 0$ the expected loss is: $(y-u)\ \tau$
- If $y - u < 0$ the expected loss is: $(y-u)(\ \tau - 1)$
- In our example the expected loss is:

$$\rho_\tau(u) = [\tfrac{1}{9}\sum_{y_i \geqq u}(y_i - u)]\tau + [\tfrac{1}{9}\sum_{y_i < u}(y_i - u)](\ \tau - 1)$$

The Loss Function: Example

- If we want to find the median $\tau = 0.5$:

$$\rho_\tau(u) = [\tfrac{1}{9}\sum_{y_i \geqq u}(y_i - u)]0.5 + [\tfrac{1}{9}\sum_{y_i < u}(y_i - u)](\ 0.5 - 1)$$

- Expected loss if u = 3:

$$\rho_\tau(3) = [\tfrac{0.5}{9}\sum_{i=3}^{9}(i_i - 3)] - [\tfrac{0.5}{9}\sum_{i=1}^{2}(i - 3)] =$$
$$\tfrac{0.5}{9}[0+1+2+3+4+5+6] - \tfrac{0.5}{9}[-2-1] = \tfrac{0.5}{9}[24]$$

- Expected loss if u = 5:

$$\rho_\tau(5) = [\tfrac{0.5}{9}\sum_{i=5}^{9}(i_i - 5)] - [\tfrac{0.5}{9}\sum_{i=1}^{4}(i - 5)] =$$
$$\tfrac{0.5}{9}[0+1+2+3+4] - \tfrac{0.5}{9}[-4-3-2-1] = \tfrac{0.5}{9}[20]$$

The Loss Function: Example

- Calculating the loss for all values of u:

u	1	2	3	4	5	6	7	8	9
$\rho_{0.5}(u) = \frac{0.5}{9} *$	36	29	24	21	20	21	24	29	36

- The expected loss is therefore minimized at 5 -> we get the median.

The Loss Function: Example bottom 1/3 decile

- We would like to find the bottom 1/3 of the distribution: $\tau = 1/3$
  The expected loss is:

$$\rho_\tau(u) = [\tfrac{1}{9} \sum_{y_i \geqq u} (y_i - u)\tfrac{1}{3}] + [\tfrac{1}{9} \sum_{y_i < u} (y_i - u)](\ \tfrac{1}{3} - 1)$$

$\longrightarrow$ Now the loss function weights positive and negative terms asymmetrically.

- Expected loss if u = 3:

$$\rho_\tau(3) = [\tfrac{1}{27} \sum_{i=3}^{9} (i_i - 3)] - [\tfrac{2}{27} \sum_{i=1}^{2} (i - 3)] =$$
$$\tfrac{1}{27}[0 + 1 + 2 + 3 + 4 + 5 + 6] - \tfrac{2}{27}[-2 - 1] = \tfrac{21}{27} + \tfrac{5}{27} = \tfrac{26}{27}$$

- Expected loss if u = 5:

$$\rho_\tau(5) = [\tfrac{1}{27} \sum_{i=5}^{9} (i_i - 5)] - [\tfrac{2}{27} \sum_{i=1}^{4} (i - 5)] =$$
$$\tfrac{1}{27}[0 + 1 + 2 + 3 + 4] - \tfrac{2}{27}[-4 - 3 - 2 - 1] = \tfrac{10}{27} + \tfrac{19}{27} = \tfrac{29}{27}$$

The Loss Function: Example

- Calculating the loss for all values of u:

u	1	2	3	4	5	6	7	8	9
$\rho_{0.33}(u) = \frac{1}{27} *$	36	30	26	27	29				

- The expected loss is therefore minimized at 3 -> we get the 33.33... percentile

Estimating Quantile Regression

- To estimate a quantile regression we substitute a linear model for $q(X_i)$

$$Q_\tau[Y_i|X_i] = \arg\min_b E[\rho_\tau(Y_i - X_i'b)]$$

- The quantile regression estimator $\widehat{\beta}_\tau$ is the regression analog of this minimization.
- Quantile regression fits a linear model to Y_i using the asymmetric loss function ρ_τ.

Quantile Regression - An Example: Returns to Education

- In one of the previous lectures you may have looked at Angrist and Krueger's (1992) paper on the returns to education.
- Rising wage inequality has prompted labour economists to investigate whether returns to education changed differently for people at different deciles of the wage distribution.
- Angrist, Chernozhukov, and Fernandez-Val (2006) show some evidence on this in their 2006 Econometrica paper.

Quantile Regression Results

Adapted from Angrist, Chernozhukov, and Fernandez-Val (2006) as reported in Angrist and Pischke (2009)

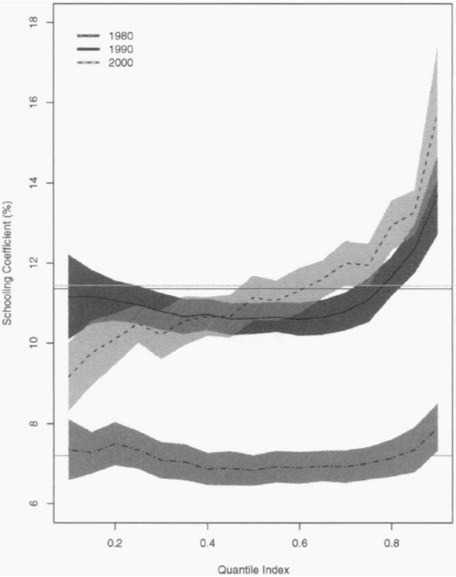
Quantile regression coefficients for schooling in the 1980, 1990, and 2000 censuses

		Desc. Stats.		Quantile Regression Estimates					OLS Estimate	
Census	Obs.	Mean	SD	0.1	0.25	0.5	0.75	0.9	Coeff.	Root M
1980	65,023	6.4	.67	.074	.074	.068	.070	.079	.072	.63
				(.002)	(.001)	(.001)	(.001)	(.001)	(.001)	
1990	86,785	6.5	.69	.112	.110	.106	.111	.137	.114	.64
				(.003)	(.001)	(.001)	(.001)	(.003)	(.001)	
2000	97,397	6.5	.75	.092	.105	.111	.120	.157	.114	.69
				(.002)	(.001)	(.001)	(.001)	(.004)	(.001)	

Quantile Regression Results

- 1980:
  - an additional year of education raises mean wages by 7.2 percent.
  - an additional year of education raises *median* wages by 0.68 percent.
  - slightly higher effects at lower and upper quantiles.
- 1990:
  - Returns go up at all quantiles and returns remain fairly similar at all quantiles.
- 2000:
  - returns start to diverge:
  - an additional year of education raises wages in the *lowest decile* by 9.2 percent.
  - an additional year of education raises wages at the *median* by 11.1 percent.
  - an additional year of education raises wages in the *highest decile* by 15.7 percent.

Graphical Illustration



Censored Quantile Regression

- Quantile regression can also be a useful tool to investigate censored data.
- Many datasets include censored data (e.g. earnings are topcoded).
- Suppose the data takes the form:

$$Y_{i,obs} = Y_i * I[Y_i < c] + c * I[Y_i \geqq c]$$

- Y_i is the variable that one would like to see but one only observes $Y_{i,obs}$ which will be equal to c where the real Y_i is greater than c.
- Quantile regression can be used to estimate the effect of covariates on conditional quantiles that are below the censoring point (if the data is censored from above).
- If earnings were censored above the median we could still estimate effects at the median.

Censored Quantile Regression - Estimation

- Powell (1986) proposed the quantile estimator:

$$Q_\tau[Y_i|X_i] = \min(c, X_i'\beta_\tau^c)$$

- The parameter vector β_τ^c solves:

$$\beta_\tau^c = \arg\min_b E\{\ I[X_i'b < c] * \rho_\tau(Y_i - X_i'b)\ \}$$

- We solve the quantile regression minimization problem for values of X_i such that $X_i'b < c$.
- In practice we solve the sample analog of this. This is no longer a linear programming problem. There are different algorithms but Buchinsky (1994) proposes a simple iterated algorithm:
  1. First estimate β_τ^c ignoring censoring.
  2. Then find the cells with $X_i'b < c$.
  3. Re-estimate the quantile regression using only the cells identified in 2. And so on...
  4. Bootstrap standard errors.