Econ 705

Jack Porter

jrporter@ssc.wisc.edu

OH: Mondays 4–5:30

and by appt.

Han Jiang

University of Wisconsin
Department of Economics

Spring Semester 2015
Economics 705

Econometrics II

Professor Jack Porter
Social Sciences Building #6448
jrporter@ssc.wisc.edu
Office Hours: Monday 4:00–5:30 and by appointment

TA Wooyoung Kim
Social Sciences Building #6473
wkim68@wisc.edu
Office Hours will take place in Social Sciences #7218: Tuesday 4:00-5:15, Wednesdays 2:30-3:30, and by appointment

Course Description

This is the second course in a three course sequence in Masters level econometrics. We will introduce a number of core topics in econometrics, including discrete choice, instrumental variables and generalized method of moments, panel data, and time series. These topics and others will be discussed from both a theoretical and applied point of view.

Lectures and Sections

Each week there will be two lectures of 1 1/4 hours. In addition there will be a section meeting once a week in which problems sets and other issues from lecture will be discussed.

Books

The textbook for the course is:
GREENE, WILLIAM, *Econometrics Analysis*, 5th edition or later, Prentice Hall.

Other helpful references are:
GOLDBERGER, A., *A Course in Econometrics*
WOOLDRIDGE, J., *Econometric Analysis of Cross Section and Panel Data*
STOCK, J. AND M. WATSON, *Introduction to Econometrics*

Problems Sets and Exams

Problem sets are an important part of the class where students will conduct their own empirical analysis using the econometric techniques discussed in class. We will be using an econometrics software package called STATA. Problem sets will be discussed in the section meetings:

4-6 problems

Semester grades will be based on a midterm exam (30%), a final exam in May (50%), and the problem sets (20%).

Wednesday, March 11

Course Outline (Parenthetical chapters refer to Greene)

- Efficiency and Robust Variance Estimation (Ch. 10, 11, 12)
  - Levitt, S. (1997), "Using Electoral Cycles in Policy Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, 87(3), 270-290
  - McCrary, J. (2002), "Using Electoral Cycles in Policy Hiring to Estimate the Effect of Police on Crime: Comment," *American Economic Review*, 92(4), 1236-1243
  - Newey, W., and West, K. (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703-708
  - Jansson, M. (2004), "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72(3), 937-946
- Weak Instruments (Ch. 5)
  - Angrist, J., and Krueger, A. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106, 979-1014
  - Bound, J., Jaeger, D., and Baker, R. (1995), "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association*, 90, 442-450
- Panel Data Models (Ch. 13)
  - Ashenfelter, O., and Krueger, A. (1994), "Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review*, 84(5), 1157-1173
  - Bonjour, D., Cherkas, L., Haskel, J., Hawkes, D., and Spector, T. (2003), "Returns to Education: Evidence from U.K. Twins," *American Economic Review*, 93(5), 1799-1812
- Maximum Likelihood Estimation (Ch. 17)
  - Bootstrap (Ch. E.5)
    - Horowitz, J. (2003), "The Bootstrap in Econometrics," *Statistical Science*, 18(2), 211-218

- Discrete Choice Models (Ch. 21)
  - McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. P. Zarembka, Academic Press; New York, 105-142
  - Manski, C. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 70(2), 519-546
  - Nevo, A. (2000), "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9(4), 513-548

- Quantile Regression (Ch. 16)
  - Koenker, R., and Hallock, K. (2001), "Quantile Regression," *Journal of Economic Perspectives*, 15, 143-156
  - Arias, O., Hallock, K., and Sosa, W. (2001), "Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression using Twins Data," *Empirical Economics*, 26(1), 7-40

- Time Series/ VARs (Ch. 19, 20)
  - Sims, C. (1980), "Macroeconomics and Reality," *Econometrica*, 48(1), 1-48
  - Stock, J., and Watson, M. (2010), *Introduction to Econometrics*, Chap. 16.1, 638-642

Recording.

Audio and/or video recording of the class is prohibited.

Students with Disabilities

Please contact Prof. Porter during the first two weeks of the semester if you have a documented requirement for accommodation to obtain equal access to this class or to any assignment in this class.

Students with Religious Observance Conflicts

Please contact Prof. Porter during the first two weeks of semester if you have a religious observance conflict on certain dates this semester for which you will need relief.

Grievance Procedure

The Department of Economics has developed a grievance procedure through which you may register comments or complaints about a course, an instructor, or a teaching assistant. The Department continues to provide a course evaluation each semester in every class. If you wish to make anonymous complaints to an instructor or teaching assistant, the appropriate vehicle is the course evaluation. If you have a disagreement with an instructor or a teaching assistant, we strongly encourage you to try to resolve the dispute with him or her directly. The grievance procedure is designed for situations where neither of these channels is appropriate.

If you wish to file a grievance, you should go to Social Science Room 7238 and request a Course Comment Sheet. When completing the comment sheet, you will need to provide a detailed statement that describes what aspects of the course you find unsatisfactory. You will need to sign the sheet and provide your student identification number, your address, and a phone where you can be reached. The Department plans to investigate comments fully and will respond in writing to complaints.

Your name, address, phone number, and student ID number will not be revealed to the instructor or teaching assistant involved and will be treated as confidential. The Department needs this information, because it may become necessary for a commenting student to have a meeting with the department chair or a nominee to gather additional information. A name and address are necessary for providing a written response.

Academic Misconduct

Academic Integrity is critical to maintaining fair and knowledge based learning at UW Madison. Academic dishonesty is a serious violation: it undermines the bonds of trust and honesty between members of our academic community, degrades the value of your degree and defrauds those who may eventually depend upon your knowledge and integrity.

Examples of academic misconduct include, but are not limited to: cheating on an examination (copying from another student's paper, referring to materials on the exam other than those explicitly permitted, continuing to work on an exam after the time has expired, turning in an exam for regrading after making changes to the exam), copying the homework of someone else, submitting for credit work done by someone else, stealing examinations or course materials, tampering with the grade records or with another student's work, or knowingly and intentionally assisting another student in any of the above. Students are reminded that online sources, including anonymous or unattributed ones like Wikipedia, still need to be cited like any other source; and copying from any source without attribution is considered plagiarism.

The Dept. of Economics will deal with these offenses harshly following UWS14 procedures (http://students.wisc.edu/saja/misconduct/UWS14.html): 1. The penalty for misconduct in most cases will be removal from the course and a failing grade, 2. The department will inform the Dean of Students as required and additional sanctions may be applied. 3. The department will keep an internal record of misconduct incidents. This information will be made available to teaching faculty writing recommendation letters and to admission offices of the School of Business and Engineering.

If you think you see incidents of misconduct, you should tell your instructor about them, in which case they will take appropriate action and protect your identity. You could also choose to contact our administrator Tammy Herbst -Koel therbst@wisc.edu) and your identity will be kept confidential.

1.

Today

Large Sample Review

OLS

Course Overview

New

OLS - Heteraskesticity

Reading

Greene

"Non-spherical Disturbances.

"Heteroskedasticity"

username bucky

password: bucky

Review

Expectation, Conditional Expectation.

Convergence in Probability in distribution.

Simplest LLN (khintchine's)

If $Z_1, Z_2, \cdots$ is iid

sequence and $E(Z_i^2) < \infty$,

then $\frac{1}{n} \sum_{i=1}^{n} Z_i \xrightarrow{P} E(Z_i)$

Simplest CLT (Lindeberg - Levy)

Suppose $Z_1, Z_2, \cdots$ is iid sequence and $E(Z_i^2) < \infty$,

then $\frac{\bar{Z}_n - E(\bar{Z}_n)}{\sqrt{Var(\bar{Z}_n)}} = \sqrt{n} \frac{\bar{Z}_n - E(Z_i)}{\sqrt{Var(Z_i)}}$

$(\bar{Z}_n = \frac{1}{n} \sum_{i=1}^{n} Z_i)$

$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i - E(Z_i)}{\sqrt{Var(Z_i)}} \xrightarrow{d} N(0,1)$

Properties:

$$Z_n \xrightarrow{d} Z_0 \qquad W_n \xrightarrow{p} c$$
$$W_n Z_n \xrightarrow{d} c Z_0$$

$$Z_n - W_n \xrightarrow{p} 0, \quad W_n \xrightarrow{d} W_0$$
$$\Rightarrow Z_n \xrightarrow{d} W_0$$

$g(\cdot)$ continous everywhere

$$Z_n \xrightarrow{p} Z_0 \Rightarrow g(Z_n) \xrightarrow{p} g(Z_0)$$
$$W_n \xrightarrow{d} W_0 \Rightarrow g(W_n) \xrightarrow{d} g(W_0)$$

Linear Regression Equation

$$y_i = x_i' \beta + \varepsilon_i, \quad i = 1, \dots, n$$

y_i, ε_i, scalar

x_i β $k \times 1$

(y_i, x_i) observed

ε_i unobserved

β unknown

$$e_i = y_i - x_i' b$$

Typical "OLS" Framework Assumptions:

(OLS0) (y_i, x_i) and

(OLS 1) $E(x_i x_i')$ finite

nonsingular

(OLS 2) $E(\varepsilon_i | x_i) = 0$ — implied by ~~economic~~ economic behavior

(OLS 3) $E(\varepsilon_i^2 | x_i) = \sigma^2$ Homoskedasticity

or $Var(\varepsilon_i | x_i) = \sigma^2$

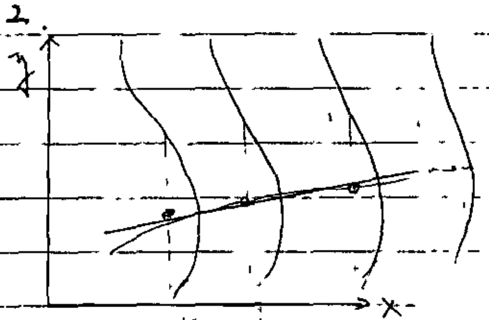
OR OLS 2 $E(y_i | x_i) = x_i' \beta$

Alternative

(OLS 2') $E(X_i \varepsilon_i) = 0$

$\underbrace{\text{(OLS 2)}}_{\text{stronger assumption.}} \Rightarrow \text{(OLS 2')}$, $\not\Leftarrow$

OLS 2.

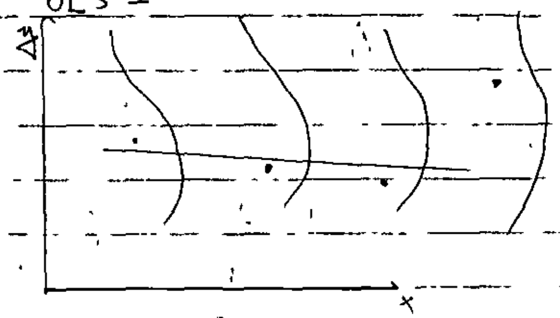


$\Downarrow$

$E(\varepsilon_i | X_i) = 0 \Rightarrow E(y_i | X_i) = X_i'\beta$

assumption $\Rightarrow$ on the line.

OLS 2'



best fitting line to this equation

$\min_{b} \left[E[y(x) - xb)^2 \right]$

$$\begin{aligned}
0 &= E(x_i \varepsilon_i) \\
&= E(x_i (y_i - x_i'\beta)) \\
&= E(x_i y_i - x_i x_i' \beta) \\
&= E(x_i y_i) - E(x_i x_i')\beta
\end{aligned}$$

$E(x_i x_i')\beta = E(x_i y_i)$

$\beta = E(x_i y_i) \cdot E(x_i x_i')^{-1}$ $\Rightarrow$ definition of β

$\rightarrow$ the slope of the best fitting lines

economic interpretation.

OLS 2' strong assumption, but good to economic interpretation.

Today
OLS properties
WLS
NEXT
IV / 2SLS
Levitt (97)

Assumptions

$y_i = x_i'\beta + \varepsilon_i \qquad i = 1 \dots n$

(OLS 0) (x_i, y_i) iid

(OLS 1) $E(x_i' x_i)$ finite nonsingular

(OLS 2) $E(\varepsilon_i | x_i) = 0$

(OLS 3) $E(\varepsilon_i^2 | x_i) = \sigma^2$ homoskedasticity

$\Rightarrow \sigma^2 = E(\varepsilon_i^2)$

(OLS 2') $E(x_i \varepsilon_i) = 0$

$= Var(\varepsilon_i | x_i)$

$= Var(y_i | x_i)$

Theorem

(a) Under OLS 0-1-2',
$\hat\beta_{OLS} \xrightarrow{p} \beta$

(b) Under OLS 0-1-2-3
$s^2 \xrightarrow{p} \sigma^2$

$\hat\beta_{OLS} = (X'X)^{-1}X'y$

$= \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i y_i$

$\xrightarrow{p} E(x_i x_i')^{-1} E(x_i y_i)$

$\underset{n \times k}{X} = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{p} E(x_i x_i')$

$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \xrightarrow{p} E(x_i x_i')^{-1}$

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i \xrightarrow{p} E(x_i y_i)$$

$$E(x_i y_i) = E(x_i(x_i'\beta + \varepsilon_i))$$
$$= E(x_i x_i')\beta + \underbrace{E(x_i\varepsilon_i)}_{(OLS\,2')\ =0}$$
$$= E(x_i x_i')\beta$$

$$\hat{\beta}_{OLS} \rightarrow \beta$$

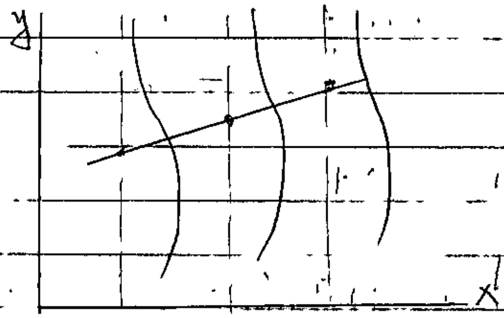
$$E(x_i\varepsilon_i) = E(E(x_i\varepsilon_i|x_i'))$$
$$= E(x_i E(\varepsilon_i|x_i)) \quad OLS\,2)$$
$$= 0$$

$$S^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$
$$\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}_{OLS}$$
$$= y_i - x_i'\beta - x_i'(\hat{\beta}_{OLS} - \beta)$$
$$= \varepsilon_i - x_i'(\hat{\beta}_{OLS} - \beta)$$

$$S^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = \frac{n}{n-k}\left[\frac{1}{n}\sum_{i=1}(\varepsilon_i - x_i'(\hat{\beta}_{OLS}-\beta))^2\right]$$
$$= \frac{n}{n-k}\left[\frac{1}{n}\sum\varepsilon_i^2 - 2\varepsilon_i x_i'(\hat{\beta}_{OLS}-\beta) + (\hat{\beta}_{OLS}-\beta)x_i x_i'(\hat{\beta}_{OLS}-\beta)\right]$$
$$= \frac{n}{n-k}\left[\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 2\frac{1}{n}\sum_{i=1}\varepsilon_i x_i'(\hat{\beta}_{OLS}-\beta) + (\hat{\beta}_{OLS}-\beta)\frac{1}{n}\sum_{i=1}x_i x_i'(\hat{\beta}_{OLS}-\beta)\right]$$

$$\downarrow \qquad \downarrow p \qquad \downarrow p \quad \downarrow p \qquad \downarrow p \quad \downarrow p$$
$$1 \qquad E(\varepsilon_i^2) \;-\; 2 \cdot 0 \cdot 0 \;+\; 0 \cdot E(x_i x_i') \cdot 0$$

$$\xrightarrow{p} 1 \cdot E(\varepsilon_i^2) = \sigma^2$$



(c) Under OLS 0-1-2-3 $\sqrt{n}(\hat{\beta}_{OLS}-\beta) \xrightarrow{d} N(0, \sigma^2 E(x_i x_i')^{-1})$

$$\sqrt{n}(\hat{\beta}_{OLS}-\beta) = \sqrt{n}\left[\left(\tfrac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\tfrac{1}{n}\sum_{i=1}^{n} x_i y_i\right) - \beta\right]$$

$$= \sqrt{n}\left[\left(\tfrac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \tfrac{1}{n}\sum_{i=1}^{n} x_i (x_i'\beta + \varepsilon_i) - \beta\right]$$

$$= \sqrt{n}\left[\underbrace{\left(\tfrac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \tfrac{1}{n}\sum_{i=1}^{n} x_i x_i'\beta}_{\beta} + \left(\tfrac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \tfrac{1}{n}\sum x_i \varepsilon_i - \beta\right]$$

$$= \underbrace{\left(\tfrac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}}_{\downarrow p \; E(x_i x_i')^{-1}} \underbrace{\tfrac{1}{\sqrt{n}}\sum x_i \varepsilon_i}_{\downarrow d \; N(0, Var(x_i\varepsilon_i))}$$

$$Var(x_i\varepsilon_i) = E(\varepsilon_i^2 x_i x_i') - \underbrace{E(\varepsilon_i x_i)}_{0} \underbrace{E(\varepsilon_i x_i')}_{0}$$

$$= E(E(\varepsilon_i^2 | x_i) x_i x_i') = \sigma^2 E(x_i x_i')$$

$$\xrightarrow{d} N(0, E(x_i x_i')^{-1} E(\varepsilon_i^2 x_i x_i') E(x_i x_i')^{-1}] \sim N(0, \sigma^2 E(x_i x_i')^{-1}]$$

(OLS 3') $E(\varepsilon_i^2 x_i x_i')$ finite

$\begin{pmatrix} \text{OLS } 1 \\ \text{OLS } 3 \end{pmatrix} \rightarrow (\text{OLS } 3')$

(c') Under OLS 0'-1-2-3'

$$\sqrt{n}(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, \underbrace{E(x_i x_i')^{-1} E(\varepsilon_i^2 x_i x_i') E(x_i x_i')^{-1}}_{\text{AVAR}})$$

(Asymptotic Variance)

homoskedasticity

$AVAR(\hat{\beta}) = \sigma^2 E(x_i x_i')^{-1}$

Heteroskedasticity

$AVAR(\hat{\beta}) = \underline{E(x_i x_i')^{-1}} \, \underline{E(\varepsilon_i^2 x_i x_i')} \, \underline{E(x_i x_i')^{-1}}$

$$\widehat{AVAR} = s^2 \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1}$$

$$\widehat{AVAR} = \left(\frac{1}{n} \sum_{i=1} x_i x_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1} \varepsilon_i^2 x_i x_i' \right) \left(\frac{1}{n} \sum_{i=1} x_i x_i' \right)^{-1}$$

what's heteroskedasticity Robust Variance Estimator (1980s)

Efficiency

Today
- Efficiency
- WLS / GLS
IV / 2SLS
(call) NEXT
Levitt
Mo

$y = x_i'\beta + \varepsilon_i$
(OLS, 0) (y_i, x_i) and
(OLS, 1) $E(x_i x_i')$ finite
nonsingular
(OLS, 2) $E(\varepsilon_i | x_i) = 0$
(OLS 2') $E(x_i \varepsilon_i) = 0$
(OLS 3) $E(\varepsilon_i^2 | x_i) = \sigma_i^2$
(OLS 3') $E(\varepsilon_i^2 | x_i x_i')$ finite

$\sqrt{n}(\hat\beta_{OLS} - \beta) \xrightarrow{d} N(0, E(x_i x_i')^{-1} E(\varepsilon_i^2 x_i x_i') E(x_i x_i')^{-1})$
AVAR$(\hat\beta_{OLS})$
under OLS3'
ie. heteroskedasticity

$\widehat{AVAR} = \left(\frac{1}{n}\sum_i x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \hat\varepsilon_i^2 x_i x_i' \left(\frac{1}{n}\sum_{i=1} x_i x_i'\right)^{-1}$
$\hat\varepsilon = y_i - x_i\hat\beta_i$

$(X'X)^{-1} X'\Omega X (X'X)^{-1}$
Var$(\varepsilon_i | X)$
Σ Var$(\varepsilon_i x_i | x_i x_i')$

W_1, W_2 indep.

$\mu = E W_1 = E W_2$ $\qquad \sigma_1^2 = Var(W_1)$

Estimate μ $\qquad \sigma_2^2 = Var(W_2)$

Find minimum Variance - unbiased estimator of the form: $\hat{\mu} = aW_1 + bW_2$

$\min Var(aW_1 + bW_2)$ $\quad$ s.t. $E[aW_1 + bW_2] = \mu$

unbiasedness:

$\mu = aEW_1 + bEW_2 = (a+b)\mu$

$\Rightarrow a+b = 1$

$\min_a V(aW_1 + (1-a)W_2)$

$$a^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$ (smaller variance, more weight)

$$1 - a^* = \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

W_2 (weight)

Regression Model

$y_i = x_i'\beta + \varepsilon_i$

$\sigma^2(x_i) = E(\varepsilon_i^2 \mid x_i)$

$$\frac{y_i}{\sigma(x_i)} = \left(\frac{x_i}{\sigma(x_i)}\right)'\beta + \frac{\varepsilon_i}{\sigma(x_i)}$$

$\tilde{y}_i = \tilde{x}_i'\beta + \tilde{\varepsilon}_i$

$$E(\tilde{\varepsilon}_i^2 \mid x_i) = E\left(\frac{\varepsilon_i^2}{\sigma^2(x_i)} \mid x_i\right) = \frac{E(\varepsilon_i^2 \mid x_i)}{\sigma^2(x_i)} = \frac{\sigma_i^2(x_i)}{\sigma^2(x_i)}$$

Weight Least Squares / Generalized Least Squares

$$\hat{\beta}_{GLS} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} = \left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{y}_i$$

$$= \left(\frac{1}{n}\sum_{i=1}^n \frac{x_ix_i'}{\sigma^2(x_i)}\right)^{-1}\frac{1}{n}\sum_{i=1}^n \frac{x_iy_i}{\sigma^2(x_i)} \leftarrow y_i = x_i'\beta + \varepsilon_i$$

$$= \beta + \left(\frac{1}{n}\sum_{i=1}^n \frac{x_ix_i'}{\sigma^2(x_i)}\right)^{-1} \cdot \frac{1}{n}\sum_{i=1}^n \frac{x_i\varepsilon_i}{\sigma^2(x_i)}$$

$$\hat\beta_{WLS} - \beta = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{x_i x_i'}{\sigma_i^2(x_i)}\right)^{-1}\frac{1}{n}\sum\frac{x_i\varepsilon_i}{\sigma_i^2(x_i)}$$

$$\downarrow p \qquad \downarrow p$$

$$E\left(\frac{x_i x_i'}{\sigma_i^2(x_i)}\right) E\left(\frac{x_i\varepsilon_i}{\sigma_i^2(x_i)}\right)$$

$$E\left(\frac{x_i E(\varepsilon_i|x_i)}{\sigma^2(x_i)}\right) = 0 \quad (OLS\ 2)$$

Theorem

(OLS 0-1-2-3')

$\Rightarrow w\hat\beta_{OLS} \xrightarrow{p} \beta$

$w\sqrt{n}\,(\hat\beta_{OLS} - \beta) \xrightarrow{d} N\left(0, E\left(\frac{x_i x_i'}{\sigma_i^2(x_i)}\right)^{-1}\right)$

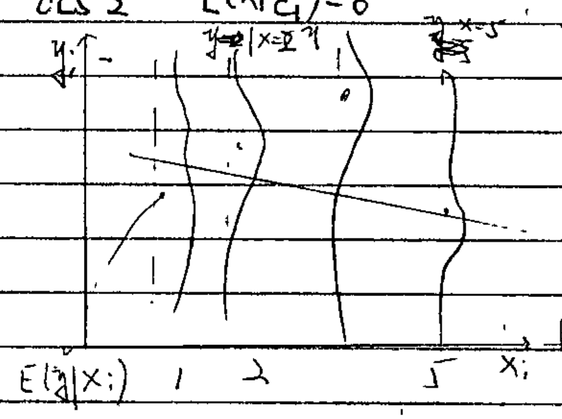
$N\left(0, V\left(\frac{x_i\varepsilon_i}{\sigma_i^2(x_i)}\right)\right)$

$$Var\left(\frac{x_i\varepsilon_i}{\sigma_i^2(x_i)}\right) = E\left(\frac{\varepsilon_i^2 x_i x_i'}{(\sigma_i^2(x_i))^2}\right)$$

$$\stackrel{(LIE)}{=} E\left(\frac{E(\varepsilon_i^2|x)\, x_i x_i'}{(\sigma_i^2(x_i))^2}\right)$$

$$= E\left(\frac{x_i x_i'}{\sigma_i^2(x_i)}\right)$$

OLS 2' $\quad E(x_i\varepsilon_i) = 0$



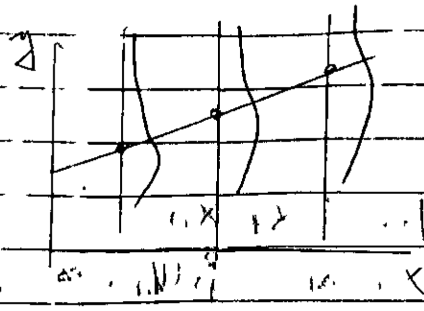
β solves $\min E[(y|x) - x'b)]$

$\beta = E(x_i x_i)^{-1} E(x_i y_i)$

$E(x_i\varepsilon_i)$ is the FOC for

"Weighted" Best Linear Prediction

$$\min_b E\left(\frac{(E(y|x) - x'\beta)^2}{\sigma^2(x)}\right]$$



assume model for σ^2

$$\sigma^2(x_i) = E(\varepsilon_i^2 | x_i)$$

$$\hat{\sigma}^2(x_i) \xrightarrow{p} \sigma^2(x_i)$$

suppose plug a wrong $\sigma^2(x_i)$

→ consistency still holds. (OLS 2 holds)

2.21

Today

- IV / 2SLS
- Levitt (1997)

Next

McCrary (2003)

- Newey-West

$$y_i = x_i'\beta + \varepsilon_i$$

(OLS 2) $E(\varepsilon_i | x_i) = 0$

(OLS 2') $E(x_i \varepsilon_i) = 0$

Suppose (OLS 2) and (OLS 2') fail!

(OLS 2') $\Rightarrow \beta = E(x_i x_i')^{-1} E(x_i y_i)$

Instrumental Variable: z_i

$E(z_i \varepsilon_i) = 0$

Levitt (97)

$C_{it} = \alpha_0 + \alpha_1 P_{it} + \alpha_2 P_{it-1} + \alpha_3 X_{it} + \varepsilon_{it}$

$P_{it} = \beta_0 + \beta_1 C_{it} + \beta_2 C_{it-1} + \beta_3 W_{it} + \mu_{it}$

- i = city
- t = time

$z_{it} = \begin{pmatrix} \text{mayoral election it} \\ \text{governor's election it} \end{pmatrix}$

(IV0) $\{y_i, x_i, z_i\}$ iid

(IV1) $E(z_i z_i')$ finite nonsingular

$E(z_i x_i')$ full-rank column

(IV2) $E(\varepsilon_i | z_i) = 0$

(IV2') $E(z_i \varepsilon_i) = 0$

(IV3) $E(\varepsilon_i^2 | z_i) = \sigma^2$

(IV3') $Var(z_i \varepsilon_i)$ infinite

$z_i y_i = z_i x_i' \beta + z_i \varepsilon_i$

$E(z_i y_i) = E(z_i x_i')\beta + \underbrace{E(z_i \varepsilon)}_{=0}$ (IV2, IV2')

$\underset{l\times 1}{E(z_i y_i)} = \underset{l\times k}{E(z_i x_i')} \underset{k\times 1}{\beta}$

$\underset{k\times 1}{x_i}$ $\underset{k\times 1}{\beta}$ $\underset{l\times 1}{z_i}$

l equs, k unknowns

l = k, "just identified"

l > k, "over identified"

$$E(y_i|z_i) = E(x_i|z)'\beta + \underbrace{E(\varepsilon_i|z_i)}_{=0\ (\text{IV2})}$$

$$E(y_i|z_i) = E(x_i|z_i)'\beta$$

OLS $E(y_i|x_i) = x_i'\beta$

Stages:

(S1) Regress x_i on z_i $\quad x_i = z_i'\pi + v_i$

$$\Rightarrow \hat{\pi} = (Z'Z)^{-1} Z'X$$

$$\Rightarrow \hat{X} = Z'\hat{\pi}$$

(S2) Regress y on $\hat{X}$ Two stage Least Squares

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$$

$$= ((Z'Z)^{-1} Z'X)^{-1} X'Z(Z'Z)^{-1} Z'y$$

Theorem

(a) IV 0–1–2' $\Rightarrow \hat{\beta}_{2sls} \xrightarrow{p} \beta$

(b) IV 0–1–2'+3' $\Rightarrow \sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0,\ [E(x_i z_i')E[z_i z_i']^{-1} E(z_i x_i')]^{-1}$

$$E(x_i z_i')E(z_i z_i')^{-1} E(\varepsilon_i^2 z_i z_i')$$

$$E(z_i z_i')^{-1} E(z_i x_i')$$

$$[E(x_i z_i')E(z_i z_i')^{-1}E(z_i x_i')]^{-1}$$

$$C_{it} = \alpha_i + \alpha_1 P_{it} + \alpha_2 P_{it-1} + \alpha_3 X_{it} + \varepsilon_{it}$$

$$C_{it} - C_{it-1} = \Delta C_{it}$$

$$= \alpha_1 \Delta P_{it} + \alpha_2 \Delta P_{it-1} + \alpha_3 \Delta X_{it} + \Delta \varepsilon_{it}$$

2.4.

TODAY

$$\hat{\beta}_{2SLS} = (X'Z\ (Z'Z)^{-1} Z'X)^{-1}\ X'Z\ (Z'Z)^{-1} Z'y$$

$$= \left[\left(\sum_{i=1}^{n} x_i z_i'\right) \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \left(\sum z_i x_i'\right) \right]^{-1}$$

$$\left(\sum_{i=1}^{n} x_i z_i'\right) \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \left(\sum z_i y_i\right)$$

$$\left(\sum z_i x_i' \beta + \sum z_i \varepsilon_i\right)$$

$$= \beta + \left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i\right) \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \left(\frac{1}{n}\sum z_i x_i'\right) \right]^{-1}$$

$$\left(\frac{1}{n}\sum x_i z_i'\right) \left(\frac{1}{n}\sum z_i z_i'\right)^{-1} \left(\frac{1}{n}\sum z_i \varepsilon_i\right)$$

$$(\hat{\beta}_{2SLS} - \beta) = \left[\left(\frac{1}{n}\sum x_i z_i'\right) \left(\frac{1}{n}\sum z_i z_i'\right)^{-1} \left(\frac{1}{n}\sum z_i x_i'\right) \right]^{-1}$$

$$\left(\frac{1}{n}\sum x_i z_i'\right) \left(\frac{1}{n}\sum z_i z_i'\right)^{-1} \frac{1}{n}\sum z_i \varepsilon_i$$

$$\downarrow p$$

$$\left[E(x_i z_i')\ E(z_i z_i')^{-1} E\ (z_i x_i) \right]^{-1}$$

$$E(x_i z_i')\ E(z_i z_i')^{-1}\ E(z_i \varepsilon_i) = 0$$

$$\downarrow p \quad \downarrow p \quad \downarrow d$$

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) = \quad N(0, V(z_i \varepsilon_i))$$

$$V(z_i \varepsilon_i) = E(\varepsilon_i^2 z_i z_i') - E(z_i \varepsilon_i)\ E(z_i \varepsilon_i)'$$

$$\hat{\beta}_{W2SLS} = \left[\left(\sum_{i=1}^{n} \frac{x_i z_i'}{w_i(z_i)}\right) \left(\sum \frac{z_i z_i'}{w_i(z_i)}\right)^{-1} \left(\sum \frac{z_i x_i'}{w_i(z_i)}\right) \right]^{-1}$$

$$\left(\sum_{i=1}^{n} \frac{x_i z_i'}{w_i(z_i)}\right) \left(\sum \frac{z_i z_i'}{w_i(z_i)}\right)^{-1} \left(\sum \frac{z_i y_i}{w_i(z_i)}\right)$$

(IV3) $E(\varepsilon_i^2 \mid z_i) = \sigma^2$

$E(\varepsilon_i^2 z_i z_i') = \sigma^2\ E(z_i z_i')$

Levitt

If (IV3) fails, define $\sigma^2(Z_i) = E(\varepsilon_i^2 | Z_i)$

$W_i(Z_i) = \sigma^2(Z_i)$ IDE A.E

Mccrary

$W_i(Z_i) = \frac{1}{\sigma_i^2(Z_i)}$

IV2 probably holding

2.9

Today

Newey West

Autocorrelation Robust

Standard Errors

NEXT

Weak IV

Angrist

Greene: simultaneous Equ chapter

(single eqn section)

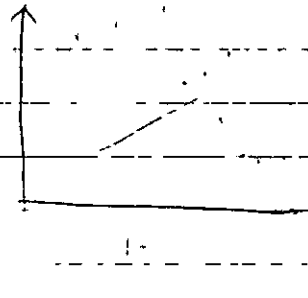
Large Sample Properties $\frac{OLS}{IV}$

$y = x_i' \beta + \varepsilon_i$

$$\hat{\beta}_{WLS} = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i'}{W(x_i)} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{x_i y_i}{W(x_i)} \right)$$

$$AVAR\left(\sqrt{n}\left(\hat{\beta}_{WLS} - \beta\right)\right) = E\left(\frac{X_i X_i'}{W(x)} \right)^{-1} E\left(\frac{\varepsilon_i^2 X_i X_i'}{W(x_i)} \right) E\left(\frac{X_i X_i'}{W(X_i)} \right)^{-1}$$

change β — on average C



even use wrong weight.

Newey - West

Definition A sequence of r.v's
$W_{-2}, W_{-1}, W_0, W_1, W_2 \dots$ is strictly stationary
if for any finite nonnegative integer M
$f_{W_t, W_{t+1}, \dots W_{t+M}}(W_t, \dots, W_{t+M}) = f_{W_s, \dots W_{s+M}}(W_s, \dots W_{s+M})$

for all t's where $f_{W_t, \dots W_{t+M}}$ is the joint density of r.v's W_t, W_{t+M}.

e.g: $f_{W_1, W_2, W_3}(W_1, W_2, W_3) = f_{W_{101}, W_{102}, W_{103}}(W_{101}, W_{102}, W_{103})$

Strong Assumption
$W_1, \dots, W_n$ iid $\Rightarrow Cov(W_1, W_2) = 0$
$W_1, \dots, W_n$ strictly stationary $\Rightarrow$ $Cov(W_1, W_2)$ may not be zero.
Weak Assumption.

$y_t = x_t'\beta + \varepsilon_t \quad t = 1 \dots T$

Assumptions:

serial correlation (SC, 0) (y_t, x_t) is strictly stationary

(SC 1) $\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \xrightarrow{p} E(x_t x_t')$ finite nonsingular

(SC 2) $E(\varepsilon_t | x_t) = 0$
$\{x_t \varepsilon_t\}$ satisfies LLN
$\frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t \xrightarrow{p} E(x_t \varepsilon_t) = 0$

(SC 3) $\{x_t \varepsilon_t\}$ satisfies CLT
$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t \varepsilon_t \xrightarrow{d} N(0, V)$
where $V = E(\varepsilon_t^2 x_t x_t') + \sum_{l=1}^{\infty}[E(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}') + E(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t')]$

Theorem
(SC 0-1-2) $\Rightarrow \hat{\beta}_{OLS} \xrightarrow{p} \beta$
(SC 0-1-2) $\Rightarrow \sqrt{n}(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, E(x_t x_t')^{-1} V E(x_t x_t')^{-1})$

$$\hat{\beta}_{OLS} - \beta = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \cdot \frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t$$

$$\downarrow p \qquad \downarrow d$$

$$E(x_t x_t')^{-1} \quad N(0, V)$$

Newey - West Method

(to estimate V)

(1) Pick a "truncation lag" G

$(G = o(T^{\alpha}), \quad 0 < \alpha < \frac{1}{4})$

(2) For $\forall\, l = 0, 1, 2, \ldots, G$ let

$$\hat{\Gamma}_l = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-l} x_t x_{t-l}' \quad \text{where} \quad \hat{\varepsilon}_t = y_t - x_t'\hat{\beta}_{OLS}$$

(3) Form $\hat{V} = \hat{\Gamma}_0 + \sum_{l=1}^{G} \left(\frac{G+1-l}{G+1}\right)\left(\hat{\Gamma}_l + \hat{\Gamma}_l'\right)$

white Estimator

TODAY

Finish Newey - West

Weak IV

NEXT

Angrist - Krueger

Bound - Jaeger Baker

$$y_t = x_t'\beta + \varepsilon_t$$

$$\sqrt{T}(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, \underbrace{E(x_t x_t')^{-1} V E(x_t x_t')^{-1}}_{AVAR(\hat{\beta}_{OLS})})$$

$AVAR(\sqrt{T}(\hat{\beta}_{OLS} - \beta))$

where

$$V = E(\varepsilon_t^2 x_t x_t') + \sum_{l=1}^{\infty}\left[E(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}') + E(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t')\right]$$

Newey - West

(1) Fix $G = O(T^{\alpha})$, $0 < \alpha < 1/4$

(2)

"Big O"

$$\left| \frac{G}{T^{\alpha}} \right| \text{ bounded as } T \to \infty$$

where does V come from?

$$V_T = Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \varepsilon_t\right)$$

$$= E\left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \varepsilon_t\right)\left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} x_s \varepsilon_s\right)'\right]$$

$$= \frac{1}{T} E\left[(x_1\varepsilon_1 + \dots + x_T \varepsilon_T)(x_1 \varepsilon_1 + \dots + x_T \varepsilon_T)'\right]$$

$$= \frac{1}{T} \sum_{t=1}^{T} E(\varepsilon_t^2 x_t x_t') + \frac{1}{T} \sum_{l=1}^{T-1} \sum_{t=l+1}^{T} \left[E(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}') + E(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t')\right]$$

$$= E(\varepsilon_t^2 x_t x_t') + \sum_{l=1}^{T-1} \frac{T-l}{T} \left[E(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}') + E(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t')\right]$$

Weak IV

$y_i = x_i \beta + \varepsilon_i$ Scalar x_i, z_i

IV z_i

(OLS 2) (OLS 2') possibly violated $E(x_i \varepsilon_i) \neq 0$

(IV 2') $E(z_i \varepsilon_i) = 0$ key: Orthogonality

(IV 1) $E(z_i z_i')$ finite nonsingular $E(z_i^2) \neq 0$

$E(z_i x_i')$ full rank $E(z_i x_i) \neq 0$

1st Stage

$x_i = z_i \pi + v_i$ $E(z_i v_i) = 0$

$$\pi = E(z_i z_i')^{-1} E(z_i x_i)$$

$$= \frac{E(z_i x_i)}{E(z_i^2)}$$

$$x_i z_i = z_i^2 \pi + z_i v_i$$

$$E(x_i z_i) = E(z_i^2)\pi + \underbrace{E(z_i v_i)}_{=0}$$

$E(x_i z_i) = E(z_i^2)\pi$

(IV1) $\Rightarrow \pi \neq 0$

Reduced Form

$y_i = x_i\beta + \varepsilon_i$

$= \underbrace{(z_i\pi + v_i)}_{=x_i}\beta + \varepsilon_i$

$y_i = z_i\overset{\alpha}{\pi\beta} + \eta_i$ where $\eta_i = \varepsilon_i + v_i\beta$

$x_i = z_i\pi + v_i$ $\qquad \alpha = \pi\beta$

π

$\beta = \frac{\alpha}{\pi} = \frac{\pi\beta}{\pi} = \beta \qquad E(z_i\eta_i) = 0.$

$\hat{\beta}_{2SLS} = [(X'Z)(Z'Z)^{-1}(Z'X)]^{-1}X'Z(Z'Z)^{-1}Z'y$

$= \frac{\sum x_i z_i (\sum z_i^2)^{-1} \sum z_i y_i}{(\sum_i x_i z_i)^2 (\sum_i z_i^2)^{-1}}$

$= \frac{\sum z_i y_i}{\sum x_i z_i} = \frac{\sum z_i (x_i\beta + \varepsilon_i)}{\sum x_i z_i}$

$= \beta + \frac{\sum z_i \varepsilon_i}{\sum x_i z_i}$ IV estimator

$\hat{\beta}_{2SLS} - \beta = \frac{\sum z_i\varepsilon_i}{\sum x_i z_i} = \frac{\frac{1}{n}\sum z_i\varepsilon_i}{\frac{1}{n}\sum z_i^2\pi + \frac{1}{n}\sum z_i v_i} \xrightarrow{p} \frac{E(z_i\varepsilon_i)}{E(z_i^2)\pi + E(z_i v_i)} \to 0$

$\hat{\beta}_{OLS} - \beta = \frac{\sum x_i\varepsilon_i}{\sum x_i^2}$ $\qquad E(v_i|\varepsilon_i) \neq 0.$

$\frac{1}{n}\sum_i z_i\varepsilon_i \xrightarrow{p} E(z_i\varepsilon_i) = 0$

$(\frac{1}{n}\sum z_i^2)\pi \xrightarrow{p} E(z_i^2)\pi$

$(\frac{1}{n}\sum_i z_i v_i) \xrightarrow{p} E(z_i v_i) = 0$

Extreme Weak IV: $\pi = 0$

$$x_i = z_i \pi + V_i$$

in this case $\hat{\beta}_{2SLS} - \beta \not\xrightarrow{p} 0$

$$\hat{\beta}_{2SLS} - \beta = \frac{\frac{1}{n}\sum_{i=1} z_i \varepsilon_i}{\frac{1}{n}\sum_{i=1} z_i V_i} \xrightarrow{d} \text{Cauchy}$$

(IV1) $\Rightarrow \pi \neq 0$ $\quad \beta = \frac{\alpha}{\pi}$

$\pi = 0$, (IV.1) failure

β not identified

2.16th

Today & Next

Weak IV

Anderson - Rubin Test

Angrist - Krueger

Bound et al

$$\tilde{y}_i = \alpha + \tilde{x}_i \beta + \tilde{\varepsilon}_i \qquad \tilde{x}_i \ \tilde{z}_i \text{ scalars}$$

$$\tilde{x} = \gamma + \tilde{z}_i \pi + \tilde{V}_i \qquad y_i = \tilde{y}_i - \bar{\tilde{y}}$$

$$\bar{\tilde{x}}_i = \gamma + \bar{\tilde{z}}_i \pi + \bar{\tilde{v}}_i \qquad x_i = \tilde{x}_i - \bar{\tilde{x}}_i$$

$$\tilde{x}_i - \bar{\tilde{x}}_i = (\tilde{z}_i - \bar{\tilde{z}}_i)\pi + \hat{V}_i - \bar{\tilde{V}} \qquad y_i = x_i\beta + \varepsilon_i$$

$$\ddot{x}_i \qquad \ddot{z}_i \ \pi + \ddot{V}_i \qquad x_i = z_i\pi + V_i$$

$$\hat{\beta}_{2SLS} = \frac{\frac{1}{n}\sum_{i=1} z_i y_i}{\frac{1}{n}\sum_{i=1} z_i x_i} = \frac{\frac{1}{n}\sum_{i=1} z_i(x_i\beta + \varepsilon_i)}{\frac{1}{n}\sum_{i=1} z_i x_i}$$

$$= \beta + \frac{\frac{1}{n}\sum_{i=1} z_i \varepsilon_i}{\frac{1}{n}\sum_{i=1} z_i x_i}$$

$$\beta_{OLS} = \beta + \frac{\frac{1}{n}\sum x_i \varepsilon_i}{\frac{1}{n}\sum_i x_i^2}$$

correlation

$$\frac{\hat{\beta}_{2SLS} - \beta}{\beta_{OLS} - \beta} = \frac{\left(\frac{\frac{1}{n}\sum z_i \varepsilon_i}{\sqrt{\frac{1}{n}\sum \varepsilon_i^2}\sqrt{\frac{1}{n}\sum z_i^2}}\right)\left(\frac{\sqrt{\frac{1}{n}\sum x_i^2}\ \sqrt{\frac{1}{n}\sum \varepsilon_i^2}}{\frac{1}{n}\sum x_i \varepsilon_i}\right)}{\left(\frac{\frac{1}{n}\sum z_i x_i}{\sqrt{\frac{1}{n}\sum x_i^2}\sqrt{\frac{1}{n}\sum z_i^2}}\right)}$$

$$= \frac{\hat{\rho}_{z\varepsilon}}{\hat{\rho}_{x\varepsilon},\ \hat{\rho}_{zx}} = \frac{\overset{\to 0}{\hat{\rho}_{z\varepsilon}} \ / \ \overset{\to \neq 0}{\hat{\rho}_{x\varepsilon}}}{\underset{\text{by IV} \neq 0}{\hat{\rho}_{xz}}} \to 0.$$

Estimator

Is IV strong or weak?

Stock-Yogo (2005)

"Rule of Thumb" : 1st stage $F \gtrless 10$? (> strong, < weak)

Alternative weak IV Estimator

Fuller, LIML : estimator

(see Stock-Yogo-Wright) 2002

Confidential Intervals

Inference. $y_i = x_i\beta + \varepsilon_i$

Testing / confidence Intervals

$H_0: \beta = \beta_0$ vs

$H_1: \beta \neq \beta_0$ Test (β_0) < accept / reject. size 5%

95% $CI = \{\beta_0.\ \text{Test}(\beta_0) \to \text{accept}\}$

Test $\Longleftrightarrow$ CI

Goal: Construct Test robust to instrument weakness.

$y_i = x_i'\beta + w_i'\gamma + \varepsilon_i$

$x_i = z_i'\pi + w_i'\delta + v_i$

Regressors: a) exogenous W_i $E(w_i\varepsilon_i) = 0$

endogenous X_i $E(x_i\varepsilon_i) \neq 0$

Instruments: $Z_i, (W_i)$ $E(z_i\varepsilon_i) = 0.$

z_i :
exclusion restriction $\varepsilon_i = z_i'\underbrace{\lambda}_{=0} + \eta$

$\therefore$ Z_i is exclued from equotation 1.

Anderson - Rubin :

$H_0 : \beta = \beta_0$ vs $H_1 : \beta \neq \beta_0$

$\underbrace{y_i - x_i'\beta_0}_{\tilde{y}_i} = z_i'\lambda + w_i'\gamma + \varepsilon_i$ (*)

$\beta = \beta_0 \Rightarrow \lambda = 0$?

$y_i - x_i'\beta_0 = x_i'\beta - x_i'\beta_0 + w_i'\gamma + \varepsilon_i$

Perform test of $\lambda = 0$ in equation (*)

good properting : Not dependent on Wedkness of IV

Kleibergen (2002)
Score Test

Moreira (2003) :
Conditional Likelihood Ratio Test

"Mincer" Regression
(Human Capital Wage Regression)

$\log \text{earnings} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exp} + \cdots + \varepsilon$ endogeneity ?

$\varepsilon = A\gamma + \mu$
ability

① ability is unobservable

② educ. is not homogenous among people β

data set : measure of ability as intelligence such as test. (SAT, GRE)

Come up with an $\underset{?}{\underbrace{IV}}$, → endogeneity

variation from state to state

⇒ date of going to school.

Angrist - Krueger

$\ln \text{earnings} = \alpha + \beta \text{educ} + \gamma X_i + \varepsilon$

(Mincer Regression)

$\varepsilon = A\delta + \mu$

(Ability)

IV QOB (Quarter of Birth)

Allen: Born Dec 31, 1946

Jan 1, 1953. 6 years old (16 years old, 11th grade)

Bob: Born, Jan 2, 1947

Jan 1, 1953, 5 years old

Jan 1, 1954, 6 years old (16 years old, 10th grade)

1st stage: $\text{educ} = \lambda + \lambda_1 Q + \lambda_2 X_i + V_i$

Reduced Form. $\ln \text{earnings} = \xi_0 + \xi_1 Q + \xi_2 X_i + \mu_i$

Regress X on Z ⇒ $\hat{\pi}$

$x = Z\hat{\pi}$

Classical Measurement Error

True dep Var / indep Vars y_i^*, x_i^*

Observed (mismeasured) Var y_i, x_i

$y_i = y_i^* + \eta_i$

$x_i = x_i^* + v_i$

η_i indep y_i^*, x_i^*
V_i indep y_i^*, x_i^*
η, V_i indep
$E(\eta_i)=0$, $E(V_i)=0$

$\hat{\beta}_{OLS} = \left(\frac{1}{n}\sum x_i x_i'\right)^{-1} \frac{1}{n}\sum x_i y_i$ $\quad$ $y_i^* = x_i^{*\prime}\beta + \varepsilon_i$

$\downarrow p \qquad \downarrow p$ $\quad$ $E(x_i^*\varepsilon_i)=0$

$E(X_iX_i')^{-1} \quad E(X_i y_i)$

$\xrightarrow{p} \left[E(x_i^*x_i^{*\prime}) + E(V_iV_i')\right]^{-1} \cdot E(X_i^*X_i^{*\prime})\beta = \dfrac{\beta\, E(x_i^{*2})}{E(x_i^{*2}) + E(V_i^2)}$

$E(X_iX_i') = E[(X_i^* + V_i)(X_i^* + V_i)'] = E(X_i^*X_i^*) + E(V_iX_i')$ $\quad$ <1 biased

$E(X_i y_i) = E[(X_i^* + V_i)(y_i^* + \eta_i)] = E(x_i^* y_i^*) = E(X_i^*X_i^{*\prime})\beta$

2. 23th

Errors in Variables
(Measurement Error)
- Omitted Variables
- Panel Data - Fixed Effects
- Ashenfelter - Krueger

Classical Measurement Error
True y_i^*, x_i^*
Observed y_i, x_i
Regression $y_i^* = x_i^{*\prime}\beta + \varepsilon_i$
$E(x_i^*x_i^{*\prime})$ finite nonsing
$E(X_i^*\varepsilon_i) = 0$

$x_i = x_i^* + V_i$
$y_i = y_i^* + \eta_i$
η_i indep x_i^*, y_i^*, V_i
V_i indep X_i^*, y_i^*
$E(\eta_i) = 0 \quad E(V_i) = 0$

$$\hat{\beta}_{OLS} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

$$\xrightarrow{p} [E(x_i^* x_i^{*\prime}) + E(V_i V_i')]^{-1} E(x_i^* x_i^{*\prime})\beta$$

Scalar Case

$$\hat{\beta}_{OLS} \xrightarrow{p} \frac{E(x_i^{*2})}{E(x_i^{*2}) + E(v^2)} \beta$$

OR with constant term

$$\hat{\beta}_{OLS} \xrightarrow{p} \frac{\sigma_{x_i^*}^2}{\sigma_{x_i^*}^2 + \sigma_v^2} \beta$$

$$\begin{aligned} y_i &= y_i^* + \eta_i \\ &= x_i^{*\prime}\beta + \varepsilon_i + \eta_i \\ &= x_i'\beta + \varepsilon_i + \eta_i - V_i'\beta \end{aligned}$$

$$y_i = x_i'\beta + \mu$$

$$E(x_i\mu_i) = E[(x_i^* + V_i)(\underbrace{\varepsilon_i + \eta_i - V_i'\beta}_{\mu})']$$

$$= \underbrace{E(x_i^*\varepsilon_i)}_{0} + \underbrace{E(x_i^*\eta_i)}_{0} - \underbrace{E(x_i^* V_i')}_{0}\beta + \underbrace{E(V_i\varepsilon_i)}_{0} + \underbrace{E(V_i\eta_i)}_{0} - E(V_iV_i')\beta$$

$$E(x_i\mu_i) = -E(V_iV_i')\beta \neq 0 \text{ (in general)}$$

Some Solutions:

① If you have information on the signal-noise ratio

② Second measurement of x_i^*

$$z_i = x_i^* + \tilde{\varepsilon}_i \quad \text{Instrument Variables}$$

IV 2 $E(z_i\mu_i) = 0$?

$E(z_i x_i')$ full col rank

☆ If $\tilde{\varepsilon}_i$ indep of $x_i^*, y_i^*, \eta_i, \varepsilon_i, V_i$ and $E(\tilde{\varepsilon}_i) = 0$

then $E(z_i\mu_i) = 0$

$$E(z_i x_i') = E\left[(x_i^* + \varepsilon_i)(x_i^* + v_i)' \right]$$
$$= E(x_i^* x_i^{*\prime}) (x_i^{*\prime})$$

Omitted Variables

$$y_i = x_i'\beta + \underbrace{w_i'\gamma + \varepsilon_i}_{u_i}$$
$$E(x_i \varepsilon_i) = 0$$
$$E(w_i \varepsilon_i) = 0$$

w_i unobserved

$$y_i = x_i'\beta + u_i$$

$E(x_i u_i) \neq 0$ in general

$$E\left[x_i (w_i'\gamma + \varepsilon_i) \right]$$
$$= E(x_i w_i')\gamma + \overset{0}{\cancel{E(x_i \varepsilon_i)}}$$
$$= E(x_i w_i')\gamma$$

$$\hat{\beta}_{OLS} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\frac{1}{n} \sum x_i y_i \right)$$
$$\downarrow p \qquad \qquad \downarrow p$$
$$E(x_i x_i') \qquad E(x_i y_i)$$
$$= E(x_i x_i')^{-1} E\left(x_i (x_i'\beta + w_i'\gamma + \varepsilon_i) \right)$$
$$= E(x_i x_i')^{-1} \left[E(x_i x_i')\beta + E(x_i w_i')\gamma + \cancel{E(x_i \varepsilon_i)} \right]$$
$$= \beta + \underbrace{E(x_i x_i')^{-1} E(x_i w_i')\gamma}_{\text{OVB (Omitted Variables Bias)}}$$

OVB disappears if

① $\gamma = 0$

OR

② $E(x_i w_i') = 0$

$$\ln wage_{it} = \beta_0 + \beta_1 educ_{it} + \beta_2 exp_{it} + A_i\gamma + \varepsilon_{it} \quad ①$$
$$\ln wage_{is} = \beta_0 + \beta_1 educ_{is} + \beta_2 exp_{is} + A_i\gamma + \varepsilon_{is} \quad ②$$

Suppose individual
at times s and t
(panel data)

Observe wage, educ, exp
Do not observe ability

① − ② $\ln wage_{it} - \ln wage_{is} = \beta_1 (educ_{it} - educ_{is}) + \beta_2 (exp_{it} - exp_{is}) + (\varepsilon_{it} - \varepsilon_{is})$

Ashenfelter − Krueger (94)
Objective = Estimate returns to education
Twins Data

Survey: wage, educ, exp, parent's educ, twin's educ.

Fixed Effects: n individuals $i = 1, \dots, n$
+ time

Suppose

$$E[\ln wage_{it} \mid edu_{it}, exp_{it}, (A_i)] = \beta_0 + \beta_1 ed_{it} + \beta_2 exp_{it} + \gamma A_i$$

$$= \beta_1 ed_{it} + \beta_2 exp_{it} + \delta_i \quad \text{where } \delta_i = \beta_0 + \gamma A_i$$

$$= \beta_1 ed_{it} + \beta_2 exp_{it} + 0 \cdot \delta_1 + 0 \cdot \delta_2 + \dots + 1 \cdot \delta_i + \dots + 0 \cdot \delta_n$$

$$= \beta_1 ed_{it} + \beta_2 exp_{it} + d_{1,it} \delta_1 + d_{2,it} \delta_2 + \dots + d_{i,it} \delta_i + \dots + d_{n,it} \delta_n$$

where $d_{j,it} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$

use dummy variable to catch / fix the effect of δ

$$y = X\beta + D\delta + \varepsilon_i$$

$$\begin{pmatrix} \hat{\beta}_{FE} \\ \hat{\delta}_{FE} \end{pmatrix} = \left([X \; D]' [X \; D] \right)^{-1} [X \; D]' y$$

TODAY & NEXT

Panel

Fixed Effect / Within

Random Effect

Dynamics

Ash Kraeger

constant Term ~~α~~ in OLS

$$y_i = \alpha + X_i'\beta + \varepsilon_i$$

$\hat{\alpha}_{OLS}, \hat{\beta}_{OLS}$ solves $\min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - \alpha - X_i'\beta)^2$

FOC $\quad 0 = -2\sum (y_i - \hat{\alpha} - x_i'\hat{\beta})^2 \qquad \hat{\alpha}$

$\quad 0 = -2 \sum x_i (y_i - \hat{\alpha} - x_i'\hat{\beta}) \qquad \hat{\beta}$

$$\hat{\alpha} = \frac{1}{n}\sum (y_i - x_i'\hat{\beta}) = \bar{y} - \bar{x}_i'\hat{\beta}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1} y_i \qquad \bar{x} = \frac{1}{n}\sum_{i=1} x_i$

$$0 = \sum_i x_i \left[y_i - (\bar{y} - \bar{x}\hat{\beta}) - x_i\hat{\beta} \right]$$

$$= \sum_i x_i \left[(y_i - \bar{y}) - (x_i - \bar{x})'\hat{\beta} \right]$$

$$= \sum_i (x_i - \bar{x}) \left[(y_i - \bar{y}) - (x_i - \bar{x})'\hat{\beta} \right]$$

$$\beta_{OLS} = \left[\sum (x_i - \bar{x})(x_i - \bar{x})' \right]^{-1} \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$y_i = \alpha + x_i'\beta + \varepsilon_i \qquad (*)$$

$$\bar{y} = \frac{1}{n}\sum y_i$$

$$= \frac{1}{n}\sum (\alpha + x_i'\beta + \varepsilon_i)$$

$$= \alpha + \bar{x}'\beta + \bar{\varepsilon} \qquad (**)$$

$$y_i - \bar{y} = (x_i - \bar{x}_i)'\beta + (\varepsilon_i - \bar{\varepsilon})$$

panel

$y = \ln wage \quad X = (ed \quad exp)$

$D = [d_1, \dots d_n]$

$y = X\beta + D\delta + \varepsilon$

$y_{it} = x_{it}'\beta + d_{1it}\delta_1 + \dots + d_{nit}\delta_n + \varepsilon_{it}$

$$\underset{nT\times 1}{y} = \begin{pmatrix} y_{11} \\ y_{1t} \\ y_{21} \\ \vdots \\ y_{2t} \\ \vdots \\ y_{nt} \end{pmatrix} \quad \underset{nT\times 2}{X} = \begin{bmatrix} x_{11}' \\ x_{1t}' \\ x_{21}' \\ \vdots \\ x_{2t}' \\ \vdots \\ x_{nt}' \end{bmatrix} \quad D = [d_1, \dots, d_n] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & & \\ 0 & 1 & & \\ \vdots & \vdots & & \\ & 1 & & \\ & 0 & \dots & 0 \\ 0 & 0 & & 1 \\ 0 & 0 & & 1 \end{bmatrix}$$

$d_{j,it} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otw} \end{cases}$

$\delta_i = \alpha + A_i\gamma \qquad nT > n$

$$\begin{pmatrix} \hat\beta_{FE} \\ \hat\delta_{FE} \end{pmatrix} = \left([X \; D]'[X \; D] \right)^{-1} [X \; D]' y$$

$(\hat\beta_{FE}, \hat\delta_{FE})$ solve $\min_{\beta,\delta} \sum_i \sum_t (y_{it} - x_{it}'\beta - d_{1it}\delta_1 - \dots - d_{nit}\delta_n)^2$

equation $FOC(\beta)\; 0 = -2 \sum_i \sum_t x_{it}(y_{it} - x_{it}'\beta - d_{1it}\delta_1 - \dots - d_{nit}\delta_n)$

$$0 = -2 \sum_i \sum_t d_{1it}(y_{it} - x_{it}'\beta - d_{1it}\delta_1 - \dots d_{nit}\delta_n)$$
$$= -2 \sum_t (y_{1t} - x_{1t}'\beta - d_{1it}\delta_1 - \dots d_{nit}\delta_n)$$
$$= -2 \sum_t (y_{1t} - x_{1t}'\beta - \delta_1)$$
$$\Rightarrow \hat\delta_1 = \frac{1}{T}\sum_{t=1} (y_{1t} - x_{1t}'\hat\beta)$$
$$= \bar y_1 - \bar x_1'\hat\beta$$

$\bar y_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} \qquad \bar x_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$

In general

$\hat\delta_i^{FE} = \bar y_i - \bar x_i'\hat\beta_{FE}$

equ FOC $\rightarrow$ substitute $\hat{\delta}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}$

$$0 = -2 \sum_i \sum_t x_{it} [y_{it} - x_{it}'\beta - \hat{\delta}_i]$$

$$= -2 \sum_i \sum_t x_{it} [y_{it} - \bar{y}_i - (x_{it} - \bar{x}_i)'\hat{\beta}]$$

$$= -2 \sum_i \sum_t (x_{it} - \bar{x}_i) [(y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_i)'\hat{\beta}]$$

$$\hat{\beta}_{FE} = \left[\sum_i \sum_t (x_{it} - \bar{x})(x_{it} - \bar{x})'\right]^{-1} \sum_i \sum_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y})$$

$= \hat{\beta}_W$ (within)

$$y_{it} = x_{it}'\beta + \delta_i + \varepsilon_{it}$$

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it}$$

$$= \frac{1}{T} \sum_{t=1}^{T} (x_{it}'\beta + \delta_i + \varepsilon_{it})$$

$$\bar{y}_i = \bar{x}_i'\beta + \delta_i + \bar{\varepsilon}_i$$

$$\underbrace{y_{it} - \bar{y}}_{\tilde{y}_{it}} = \underbrace{(x_{it} - \bar{x}_i)'}_{\tilde{x}_{it}}\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

Common "small T" Panel Assumptions

(Pan 0) $\{(y_{i1}, x_{i1}), \dots (y_{iT}, x_{iT})\}$ iid across i.

(Pan 1) Square matrices are invertible
expectations exist

(Pan 2) $E(\varepsilon_{it} \mid x_{i1}, \dots, x_{iT}, \delta_i) = 0$ ☆

"Strict Exogenity"

restricted, restricted

$$y_{it} = x_{it}'\beta + \delta_i + \varepsilon_{it} \qquad x_{is}$$

unrestricted, restricted.

under Pan2.

$$\hat{\beta}_{FE} - \beta = \left[\sum_i \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\right]^{-1} \sum_i \sum_t (x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)$$

TODAY

Panel

Ashenfelter - Krueger

Between vs

Next

Time Dummies

Dynamic Panel

Max Likelihood

Next WEEK

Mon - Midterm Review

Wed - Midterm (March 11)

Ingraham Room 19

Twin / pair, $i = 1, \ldots, n$

Twin within pair $k = 1, 2$

(*) $y_{ki} = \beta S_{ki} + \gamma Z_{ki} + \delta_i + \varepsilon_{ki}$

y_{ki}: log earning; S_{ki}: schooling of twin k in pair i; Z_{ki}: other variables; δ_i: twin effect (unobservable ability)

$E(\varepsilon_{ki} \mid S_{1i}, S_{2i}, \delta_i, Z_{1i}, Z_{2i}) = 0$ (Strict exogeneity)

S_{ki} = true schooling twin k

S_{ki}^j = schooling for twin k reported by twin j

$S_{ki}^j = S_{ki} + V_{ki}^j$

V_{ki}^j: measurement error

Within transformation

$$Z_{1i} - \bar{Z}_i = Z_{1i} - \frac{Z_{1i} + Z_{2i}}{2}$$

$$\frac{Z_{1i} - Z_{2i}}{2}$$

(Another representation of first difference / which is also the same as FE)

OLS Regress $(y_{1i} - y_{2i})$ on $(S_{1i} - S_{2i})$, (Z_{1i}, Z_{2i})

$$\downarrow$$

$$(S^1_{1i} - S^2_{2i})$$

$y_{1i} - y_{2i}$ on $(S^1_{1i} - S^2_{2i})$ $(Z_{1i} - Z_{2i})$

IVi $(S^2_{1i} - S^1_{2i})$

Measurement error thends to shrink coefficient estimates to 0, But in this paper $\hat{\beta}_{IV} > \hat{\beta}_{OLS}$

by (*) $y_{1i} - y_{2i} = \beta(S_{1i} - S_{2i}) + \gamma(Z_{1i} - Z_{2i}) + \varepsilon_{1i} - \varepsilon_{2i}$

$$y_{1i} - y_{2i} = \beta(S^1_{1i} - S^2_{2i}) + \gamma(Z_{1i} - Z_{2i}) + [(\varepsilon_{1i} - \varepsilon_{2i}) - \beta(V^1_{1i} - V^2_{2i})]$$

IVi $(S^2_{1i} - S^1_{2i})$

For 2SLS, $E(X'u) = 0$.

(IV 2)

$$0 = E[(S^2_{1i} - S^1_{2i})\{(\varepsilon_{1i} - \varepsilon_{2i}) - \beta(V^1_{1i} - V^2_{2i})\}]$$

$$= E\{(S^*_{1i} - S^*_{2i})\}[(\varepsilon_{1i} - \varepsilon_{2i}) - \beta(V^1_{1i} - V^2_{2i})]$$

$$+ E[(V^2_{2i} - V^1_{1i})\}\{(\varepsilon_{1i} - \varepsilon_{2i}) - \beta(V^1_{1i} - V^2_{2i})\}]$$

V^j_{ki} indep. S^*_{ki}, $E(V^j_{ki}) = 0$

V^j_{ki} indep ε_{ki}

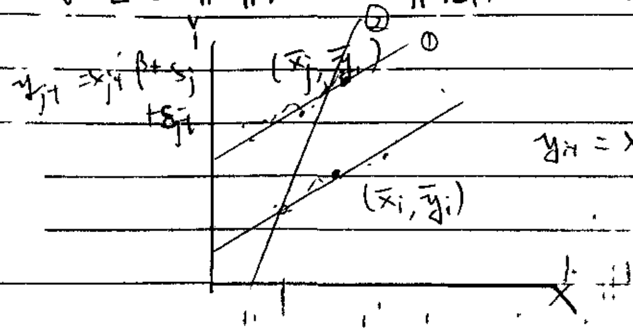
$$0 = E(V^2_{1i} V^1_{1i}) = E(V^1_{1i} V^2_{2i}) = E(V^1_{2i} V^1_{1i}) = E(V^1_{2i} V^2_{2i})$$

$$y_{1i} - y_{2i} = \beta(S^1_{1i} - S^1_{2i}) + \gamma(Z_{1i} - Z_{2i}) + (\varepsilon_{1i} - \varepsilon_{2i}) - \beta(V^1_{1i} - V^1_{2i})$$

IV: $(S_{1i}^2 - S_{2i}^2)$

$\cancel{E(V_{1i}^2 V_{2i}^2)}$ $\cancel{E(V_{2i}^1, V_{1i}^1)}$

$0 = E(V_{1i}^2 V_{1i}^1) = E(V_{1i}^2 V_{2i}^1) = E(V_{2i}^2 V_{2i}^1) = E(V_{2i}^2 V_{1i}^1)$



$y_{it} = x_{it}'\beta + \delta_i + \varepsilon_{it}$

Decomposition

$z_{it} = \underbrace{(z_{it} - \bar{z}_i)}_{\text{within transformation}} + \overset{\text{between}}{\bar{z}_i}$

Orthog Decomp...

$0 = \sum_{i=1} \sum_{t=1} (z_{it} - \bar{z}_i)\bar{z}_i = \sum_{i=1} \hat{\bar{z}}_i \left(\sum_{t=1} z_{it} - \bar{z}_i\right) = 0$

z_{it} $\cancel{\bar{z}_i = \frac{1}{T}\sum z_{it}}$

ToDay Panel Data.

Random effects

Fixed effect

Time Dummies

Dynamics

NEXT.

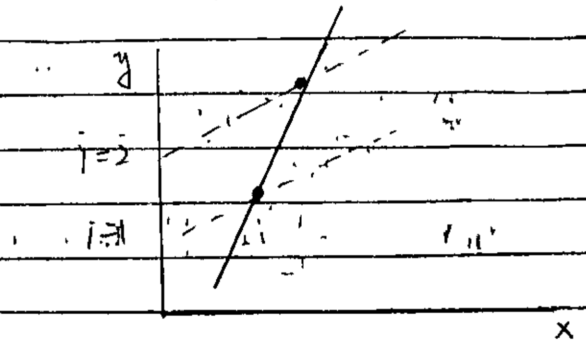
Midterm Review

Wed March 11

Midterm

Ingraham Hall 11:00 - 12:15

VI



$y_{it} = x_{it}'\beta + \delta_{it} \varepsilon_{it}$

Regression

Fixed Effects

Regress y_{it} on $x_{it}, d_{1it}, \dots d_{nit}$

Random Effects

regress y_{it} on x_{it},

Within Regression

$y_{it} - \bar{y}_i$ on $x_{it} - \bar{x}_i$

Between Regression

$\bar{y}_i$ on $\bar{x}_i$

Decomposition,

$Z_{it} = (Z_{it} - \bar{Z}_i) + \bar{Z}_i$

$\sum_{i=1}^{n} \sum_{t=1}^{T} (Z_{it} - \bar{Z}_i)(\bar{Z}_i) = 0$

$\sum_{i=1}^{n} \sum_{t=1}^{T} (Z_{it} - \bar{Z}_i) \bar{W}_i = 0$

$\hat{\beta}_{FE} = \left(\sum_i \sum_t x_{it} x_{it}'\right)^{-1} \sum_i \sum_t x_{it} y_{it}$

$= \sum_i \sum_t \Big[\underbrace{(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'}_{W_{xx}}\Big]^{-1} \sum_i \sum_t \underbrace{(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}_{W_{xy}}$

$\hat{\beta}_{Bet} = \left(\sum_i \bar{x}_i \bar{x}_i'\right)^{-1} \sum_{i=1} \bar{x}_i \bar{y}_i$

$$\hat{\beta}_{RE\cdot OLS} = \left(\sum_i \sum_t x_{it} x_{it}'\right)^{-1} \sum_i \sum_t x_{it} y_{it}$$

$$= \left\{\sum_i \sum_t [(x_{it} - \bar{x}_i) + \bar{x}_i][(x_{it} - \bar{x}_i) + \bar{x}_i]'\right\}^{-1}$$

$$\left\{\sum_i \sum_t [(x_{it} - \bar{x}_i) + \bar{x}_i][(y_{it} - \bar{y}_i) + \bar{y}_i]\right\}$$

$$= \Big\{\underbrace{\sum_i \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'}_{W_{xx}} + \underbrace{T \sum_i \bar{x}_i \bar{x}_i'}_{B_{xx}}\Big\}^{-1} \Big\{\underbrace{\sum_i \sum_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y})}_{W_{xy}}$$

$$+ \underbrace{T \sum_i \bar{x}_i \bar{y}_i}_{B_{xy}}\Big\}$$

$$\hat{\beta}_{RE\cdot OLS} = [W_{xx} + B_{xx}]^{-1}[W_{xy} + B_{xy}]$$

$$= [W_{xx} + B_{xx}]^{-1} W_{xy} + [W_{xx} + B_{xx}]^{-1} B_{xy}$$

$$= \underbrace{[W_{xx} + B_{xx}]^{-1} W_{xx}}_{A} \underbrace{W_{xx}^{-1} W_{xy}}_{\to \hat{\beta}_{FE}}$$

$$+ \underbrace{[W_{xx} + B_{xx}]^{-1} B_{xx}}_{I - A} \underbrace{B_{xx}^{-1} B_{xy}}_{= \hat{\beta}_{Bet}}$$

$$= A\, \hat{\beta}_{FE} + (I - A)\, \hat{\beta}_{Bet}$$

$$y_{it} = x_{it}'\beta + \delta_i + \varepsilon_{it}$$

FE: $E(\varepsilon_{it} \mid \delta_i, x_{i1}, \ldots, x_{iT}) = 0$ (strict exog) $\Big\} \to$ RE

$E(\delta_i \mid x_{i1}, \ldots, x_{iT}) = 0$

$q = \hat{\beta}_{FE} - \hat{\beta}_{RE}$ Hausman Test

Time Dummies

$$y_{it} = x_{it}'\beta + \delta_i + \lambda_t + \varepsilon_{it}$$

across individual $\bar{y}_t = \bar{x}_t'\beta + \bar{\delta} + \lambda_t + \bar{\varepsilon}_t$

time $\bar{y}_i = \bar{x}_i'\beta + \delta_i + \bar{\lambda} + \bar{\varepsilon}_i$

both $\bar{y} = \bar{x}'\beta + \bar{\delta} + \bar{\lambda} + \bar{\varepsilon}$

$$\bar{y}_t = \frac{1}{N} \sum_{i=1}^{N} y_{it}$$

$$\bar{y} = \frac{1}{NT} \sum_i \sum_t y_{it}$$

$$y_{it} - \bar{y}_{i} = (x_{it} - \bar{x})'\beta + \lambda_t - \bar{\lambda} + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (*)$$

$$\bar{y}_{t} - \bar{y} = (\bar{x}_t - \bar{x})'\beta + (\lambda_t - \bar{\lambda}) + (\bar{\varepsilon}_t - \bar{\varepsilon}) \quad (**)$$

$$(*) - (**) \quad y_{it} - \bar{y}_i - (\bar{y}_t - \bar{y}) = (x_{it} - \bar{x}_i - \bar{x}_t + \bar{x})'\beta$$
$$+ (\varepsilon_{it} - \bar{\varepsilon}_i - \bar{\varepsilon}_t + \bar{\varepsilon})$$

$\tilde{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$, Regression $\tilde{y}_{it}$ on $\tilde{x}_{it}$

$\tilde{x}_{it} = x_{it} - \bar{x}_i - \bar{x}_t + \bar{x}$

Dynamic Panel

$$y_{it} = \alpha \underbrace{y_{it-1}}_{\text{state dependence}} + x_{it}'\beta + \underbrace{\delta_i}_{\text{unobserved heterogeneity}} + \varepsilon_{it}$$

$$y_{it} - y_{it-1} = \alpha(y_{it-1} - y_{it-2}) + (x_{it} - x_{it-1})'\beta$$
$$+ (\varepsilon_{it} - \varepsilon_{it-1})$$

IV: y_{it-2}, y_{it-3}

3.9th 2014

WLLN, CLT

OLS 2 vs OLS 2'

(conditional expectation function vs: best linear prediction)

Consistency Assumption for OLS Normality

- Heteroskedasticity

White Robust

Variance - Covariance

Weighted LS

=

strictly stationary

Newey - West

= 2SLS

Consistency

asymptotic normality for 2SLS

Weighted 2SLS

Levitt (97) McCrary (02)

= Weak IV

OLS vs IV/2SLS

- 1st stage/ reduced form

= Anderson - Rubin Test

Angrist - Krueger (91)

- ~~[illegible]~~ Bound et al.

- classical Measurement Error

- Omitted Variables Bias

- Ashenfelter - Krueger (94)

= Fixed Effect w/ with

- Random Effects / Between

$y_i = x_i'\beta + \varepsilon$

(OLS0) (y_i, x_i) iid

(OLS1) $E(x_i x_i')$ finite nonsingular

(OLS2) $E(\varepsilon_i | x_i) = 0$. (OLS2') $E(\varepsilon_i x_i) = 0$

(OLS3) $E(\varepsilon_i^2 | x_i) = \sigma^2$, (OLS3') $E(\varepsilon_i^2 x_i x_i')$ finite

(SC0) (y_t, x_t) strictly stationary

(SC1) $\frac{1}{T}\sum_T x_t x_t' \xrightarrow{P} E(x_t x_t')$, finite nonsingular

(SC2) $E(\varepsilon_t | x_t) = 0$, $\frac{1}{T}\sum_T x_t \varepsilon_t \xrightarrow{P} 0$

(SC3) $\{x_i, \varepsilon_i\}$ satisfies CLT

$\frac{1}{\sqrt{T}}\sum_T x_t \varepsilon_t \xrightarrow{d} N(0, V)$

$V = E(\varepsilon_t^2 x_t x_t') + \sum_{l=1}^{\infty} \left(E(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}') + E(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t') \right)$

(IV0) (y_i, x_i, z_i) iid

(IV1) $E(z_i z_i')$ finite nonsingular

$E(z_i x_i')$ full column rank

(IV2) $E(\varepsilon_i | z_i) = 0$ (IV2') $E(x_i \varepsilon_i) = 0$

(IV3) $E(\varepsilon_i^2 | z_i) = \sigma_i^2$ (IV3') $E(\varepsilon_i^2 z_i z_i')$ nonsingular

panel questions

$= x_i'\beta + (\varepsilon_i - v_i'\beta)$

$y_i = x_i^{*\prime}\beta + \varepsilon_i$ a) OLS y_i on x_i (plim?)

Assume $E(\varepsilon_i | x_i^*) = 0$

Observe (y_i, x_i) where

$x_i = x_i^* + v_i$

v_i independent of x_i^* and $E(v_i) = 0$

x_i scalar x^* scalar

$E(\varepsilon_i v_i) = \gamma$

$$\hat\beta_{OLS} = \left(\sum_i x_i' x_i\right)^{-1}\left[\sum_i x_i y_i\right] = \left(\sum_i x_i x_i\right)^{-1}\sum_i x_i\,(x_i\beta + \varepsilon_i - v_i\beta)$$

$$= \left(\sum_i x_i x_i\right)^{-1}\left[\sum_i x_i x_i\,\beta + \sum_i x_i\varepsilon_i - \sum x_i v_i\beta\right]$$

$$= \beta + \left(\sum_i x_i x_i\right)^{-1}\left(\sum_i x_i\varepsilon_i - \sum_i x_i v_i\beta\right)$$

$$\hat\beta_{OLS} - \beta = \left(\frac{1}{n}\sum_i x_i x_i'\right)^{-1}\left[\frac{1}{n}\sum x_i\varepsilon_i - \frac{1}{n}\sum x_i v_i\beta\right]$$

$$= \left[\frac{1}{n}\sum_i (x_i^* + v_i)(x_i^* + v_i)\right]^{-1}$$

$$\left[\frac{1}{n}\sum x_i^*\varepsilon_i + v\varepsilon_i - x_i^* v_i\beta - v_i v_i\beta\right]$$

$$= \left[\frac{1}{n}\sum (x_i^* x_i^{*\prime} + x_i^* v_i + v_i x_i^{*\prime} + v_i v_i')\right]^{-1}$$

$$\left[\frac{1}{n}\sum_i (x_i^*\varepsilon_i + v_i\varepsilon_i - x_i^* v_i\beta - v_i v_i\beta\right]$$

$$= \left[E(x_i^{*2}) + \underbrace{E(x_i^* v_i)}_{0} + \underbrace{E(v_i x_i^*)}_{0} + E(v_i^2)\right]^{-1}\left[\underbrace{E(x_i^*\varepsilon_i)}_{0} - \underbrace{E(v_i\varepsilon_i)}_{\gamma} - \underbrace{E(x_i^* v_i\beta)}_{0} - E(v_i^2)\beta\right]$$

$$= \frac{\gamma - E(v_i^2)\beta}{E(x_i^{*2}) + E(v_i^2)}$$

3)

$$E(v_i^2)\beta = \gamma$$

$$E(v_i^2) = 2$$

$$E(\varepsilon_i v_i) = \beta/2$$

Suppose $w_i = x_i^* + \eta_i$

η_i independent of x_i^*, $E(\eta_i) = 0$

$E(\varepsilon_i\eta_i) = 0$ $\quad \dfrac{E(\eta_i v_i)}{E(x_i^{*2})} = \lambda$

$$\hat\beta_{IV} = \left(\sum_{i=1} w_i x_i\right)^{-1}\left(\sum w_i y_i\right)$$

b) plim $\hat\beta_{IV}$?

$$\hat\beta_{IV} = \left(\sum_{i=1} w_i x_i\right)^{-1}\sum_{i=1} w_i\,(x_i\beta + \varepsilon_i - v_i\beta)$$

$$= \beta + \left(\sum w_i x_i\right)^{-1}\sum_i (w_i\varepsilon_i - w_i v_i\beta)$$

9)

$$\hat\beta_{IV} - \beta = \left[\frac{1}{n}\sum (x_i^* + \eta_i)(x_i^* + v_i)\right]^{-1}\frac{1}{n}\sum\left[(x_i^* + \eta)\varepsilon_i - (x_i^* + \eta)v_i\beta\right]$$

$$= \left[\frac{1}{n}\sum_i x_i^{*2} + x_i v_i + x_i^*\eta_i + \eta_i v_i\right]^{-1}\frac{1}{n}\sum_i\left(x_i^*\varepsilon_i + \eta_i\varepsilon_i - x_i^* v_i\beta - \eta v_i\beta\right]$$

$$= \left[E(x_i^{*2}) + E(x_i^* v_i) + E(x_i^* \eta_i) + E(\eta_i v_i)\right]^{-1}$$

$$\left[E(x_i^* \varepsilon) + E(\eta_i \varepsilon_i) - E(x_i^* v_i)\beta - E(\eta_i x_i)\beta\right]$$

$$= \frac{-E(\eta_i v_i)\beta}{E(x_i^{*2}) + E(\eta_i v_i)} = \frac{-\lambda E(x_i^{*2})}{E(x_i^{*2}) + \lambda E(x_i^{*2})}\beta$$

$$= \frac{-\lambda}{1+\lambda}\beta$$

b) $\text{plim } \hat{\beta}_{IV} = \beta - \frac{\lambda}{1+\lambda}\beta = \frac{\beta + \lambda\beta}{1+\lambda} - \frac{\lambda\beta}{1+\lambda}$

$$= \frac{\beta}{1+\lambda}$$

c) $\lambda = 0 \quad \sqrt{n}(\hat{\beta}_{IV} - \beta) \quad AVAR = \frac{V}{E(x_i^*)^2}$

$$\hat{\beta}_{wIV} = \left(\frac{1}{n}\sum_i \frac{w_i x_i}{g(w_i)}\right)^{-1}\left(\frac{1}{n}\sum_i \frac{w_i y_i}{g(w_i)}\right)$$

$$E\left(\frac{w_i \varepsilon_i}{g(w_i)}\right)$$

3.16th

TODAY

- Maximum Likelihood Review
- Non linear Least Squares

NEXT:
Probit

Logit

Greene Max Likelihood chapter
(up to testing)
Models for Discrete choice
(Binary Choice Section)

Maximum likelihood

r.v. X cdf $F_x(x) \rightarrow F_x(x|\beta)$

pdf/pmf $f_x(x) \rightarrow f_x(x|\beta)$

$y = x'\beta + \varepsilon$ $\qquad f(y|x,\beta)$

$E(y|x) = x'\beta$

e.g. $x_1, \dots, x_5 \overset{iid}{\sim}$ Bernoulli(p)

$$\Pr(X = x|p) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

$f_x(x|p) = \Pr(X = x|p) = p^x(1-p)^{1-x}$ if $x \in \{0,1\}$

Suppose we know $p = 0.15$ or 0.9 and we observe Data $(x_1, x_2, \dots x_5) = (0, 1, 0, 0, 0)$

$\hat{p} = 0.15$

$\Pr(x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 0, x_5 = 0 | p = 0.15)$
$= (.85)^4 (0.15) = 0.078$

Y

$P_Y(X_1=0, X_1=1, X_2=0, \dots X_5=0 \mid p=0.9) = (0.1)^4(0.9)$
$= 0.00009$

Max Likelihood

$X_1, \dots X_n \sim f_{X_1,\dots,X_n}(X_1,\dots,X_n \mid \theta)$

joint pmf / pdf   unknown parameter

e.g. $X_1, \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$f_{X_1 \dots X_n}(X_1, \dots X_n \mid \underbrace{\mu, \sigma^2}_{\theta})$

$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

Given the data $x_1, \dots x_n$
the "likelihood" is
$\mathcal{L}(\theta \mid x_1, \dots x_n) = f_{X_1 \dots X_n}(x_1, \dots, x_n \mid \theta)$
$l(\theta \mid x_1, x_2 \dots x_n) = \ln f_{X_1 \dots X_n}(x_1, \dots x_n \mid \theta)$
✓ product → sum

θ_{ML} solves $\max_\theta \mathcal{L}(\theta \mid x_1, \dots, x_n)$

under certain conditions.
a) (Consistency) $\hat{\theta}_{ML} \xrightarrow{p} \theta_0$
b) (Asymptotic Normality)
$\sqrt{n}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N(0, J^{-1})$
where $J = E\left[\frac{\partial}{\partial\theta} \ln f_X(X \mid \theta_0)\right]\left[\frac{\partial}{\partial\theta} \ln f_X(X \mid \theta)\right]'$
(iid case)

Binary Choice
Nonlinear Least Squares
Probit
$y_i \in \{0, 1\}$

Outcomes Covariates : X_i ( ).

Assumptions

(Probit 0) (y_i, x_i) iid

(Probit 1) $E(x_i x_i')$: finite, nonsingular

(Probit 2) $E(y_i|x_i) = \Phi(x_i'\beta_0)$

$\Pr(y_i = 1|x_i)$

where $\Phi(\cdot)$ cdf of standard normal

Compare to OLS

OLS 2 $E(y_i|x_i) = x_i'\beta$ $\quad E(\varepsilon_i|x_i) = 0$ where $\varepsilon_i = y_i - x_i'\beta$

Pro 2 $E(y_i|x_i) = \Phi(x_i'\beta_i) \Rightarrow E(\varepsilon_i|x_i) = 0$ where

$y_i = \Phi(x_i'\beta) + \varepsilon_i$ $\quad \varepsilon_i = y_i - \Phi(x_i'\beta_0)$

NLLS (non linear Least squares)

$$\hat\beta_{NLLS} = \arg\min_\beta \underbrace{\frac{1}{n}\sum_{i=1}^n (y_i - \Phi(x_i'\beta))^2}_{Q_n(\beta)}$$

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^n [y_i - \Phi(x_i'\beta)]$$

$$Q_n(\beta) \xrightarrow{p} E[(y_i - \Phi(x_i'\beta))^2] \text{ for each } \beta$$

$$E[Q_n(\beta)] = E[(y_i - \Phi(x_i'\beta))^2]$$

$$= E[(y_i - \Phi(x_i'\beta_0) - \{\Phi(x_i'\beta) - \Phi(x_i'\beta_0)\})^2]$$

$$= E[(y_i - \Phi(x_i'\beta_0))^2] + E[\{\Phi(x_i'\beta) - \Phi(x_i'\beta_0)\}^2]$$

$$\underbrace{-2E[(y_i - \Phi(x_i'\beta_0))(\Phi(x_i'\beta) - \Phi(x_i'\beta_0))]}_{0}$$

$$= -2E[\underbrace{E[y_i - \Phi(x_i'\beta_0)|x_i]}_{0}(\Phi(x_i'\beta) - \Phi(x_i'\beta_0))]$$

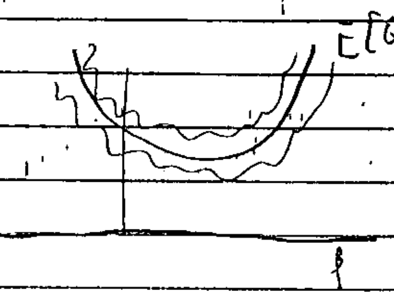
$$E[Q_n(\beta)] = E(y_i - \Phi(x_i'\beta_0))^2] + E[(\Phi(x_i'\beta) - \Phi(x_i'\beta_0))^2]$$

$\beta = \beta_0$ minimizes $E[Q_n(\beta)]$

$\hat{\beta} = \arg\min_{\beta} Q_n(\beta)$

$\beta_0 = \arg\min_{\beta} E[Q_n(\beta)]$

$Q_n(\beta) \xrightarrow{P} E[Q_n(\beta)]$



3.18th. Today
- Binary choice
- Random Utility
- Probit
- Logit

Recall $y_i \in \{0, 1\}$ covariaties X_i

$y_i = \Phi(X_i'\beta_0) + u_i \;(E(y_i|X_i) = \Phi(X_i'\beta_0)) = P_r(y_i = 1 | X_i)$

NLLS: $\hat{\beta}_{NLLN} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} (y - \Phi(x_i'\beta))^2 = Q_n(\beta)$

$\Downarrow P$

true $\beta_0 = \arg\min_{\beta} E[(y - \Phi(x_i'\beta))^2] = Q(\beta)$

Does $\hat{\beta}_{NLLS} = \arg\min Q_n(\beta)$ $\beta_0 = \arg\min_{\beta} Q(\beta)$ and $Q_n(\beta) \xrightarrow{P} Q(\beta)$, $\forall \beta$

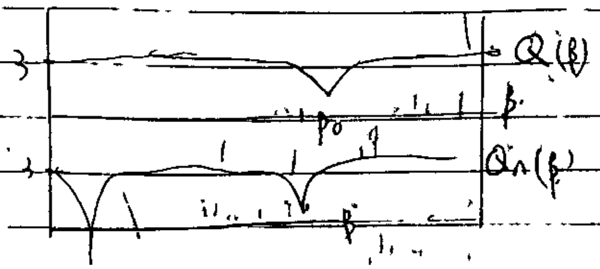
$\Rightarrow \hat{\beta}_{NLLS} \xrightarrow{P} \beta_0$???

In general, No.

"Need" uniform convergence, $\sup_{\beta} |Q_n(\beta) - Q(\beta)| \xrightarrow{P} 0$

$Q_n(\beta) \xrightarrow{P} Q(\beta)$ for each β. $\hat{\beta}_n \not\to \beta_0$

But $\sup_{\beta} |Q_n(\beta) - Q(\beta)| = 3$



Binary choice : Alternate Formulation

Random Utility $y_i^* = x_i'\beta_0 + \varepsilon_i$

Note: $y_i \neq y_i^*$ y_i^* random utility difference

Assumption : $\varepsilon_i | x_i \sim N(0,1)$

$$y_i = \begin{cases} 1 & y_i^* \geq 0 \\ 0 & y_i^* < 0 \end{cases}$$

$$\Pr(y_i = 1 | x_i) = \Pr(y_i^* \geq 0 | x_i)$$
$$= \Pr(x_i'\beta_0 + \varepsilon_i \geq 0 | x_i) = \Pr(-\varepsilon \leq x_i'\beta | x_i)$$
$$= \Phi(x_i'\beta_0)$$

$$\frac{\partial \Pr(y_i = 1 | x_i)}{\partial x_i^j} = \varphi(x_i'\beta_0) \cdot \beta_{0j}$$

"marginal effects"

$$\frac{\partial \Pr(y_i = 1 | x_i) / \partial x^j}{\partial \Pr(y_i = 1 | x_i) / \partial x^l} = \frac{\varphi(x_i'\beta_0)\beta_{0j}}{\varphi(x_i'\beta_0)\beta_{0l}} = \frac{\beta_{0j}}{\beta_{0l}}$$

3.22 wed

Today

Probit / Logit

Properties

Next

Semiparametric Binary choice

Panel Binary choice

Binary choice with endogeneity

(Probit 0) (y_i, x_i) iid

(Probit 1) $E(x_i x_i')$ finite nonsingular

(Probit 2) $E(y_i | x_i) = Pr(y_i = 1 | x_i) = \Phi(x_i' \beta_0)$

$y_i = \{0, 1\}$, x_i characteristics / covariates

$y_i = \Phi(x_i'\beta) + u_i$

$E(u_i | x_i) = 0$

Likelihood (Conditional)

$\mathcal{L}(\beta | (y_1, x_1), \ldots, (y_n, x_n)) = \prod_{i=1}^{n} \Phi(x_i'\beta)^{y_i} [1 - \Phi(x_i'\beta)]^{1-y_i}$ $y_i = 0$

$L(\beta) = \ln \mathcal{L}(\beta) = \sum_{i=1}^{n} y_i \ln \Phi(x_i'\beta) + (1 - y_i)(1 - \Phi(x_i'\beta))$

$\hat{\beta}_{Probit} = \arg\max_\beta \mathcal{L}_n(\beta) = \arg\max_\beta L_n(\beta)$

$= \arg\max_\beta \frac{1}{n} L_n(\beta)$

Note: $\frac{1}{n} L_n(\beta) \xrightarrow{P} E[y_i \ln \Phi(x_i'\beta) + (1 - y_i)(1 - \Phi(x_i'\beta))]$

$= E[\ln(\Phi(x_i'\beta)^{y_i} [1 - \Phi(x_i'\beta)]^{1-y_i})]$

$= L(\beta)$

show β_0 maximizes $L(\beta)$

$L(\beta) - L(\beta_0)$

$= E[\ln(\Phi(x_i'\beta)^{y_i} (1 - \Phi(x_i'\beta))^{1-y_i})]$

$- E[\ln(\Phi(x_i'\beta_0)^{y_i} (1 - \Phi(x_i'\beta_0))^{1-y_i})]$

$= E\left[\ln\left(\frac{\Phi(x_i'\beta)^{y_i} (1 - \Phi(x_i'\beta))^{1-y_i}}{\Phi(x_i'\beta_0)^{y_i} (1 - \Phi(x_i'\beta_0))^{1-y_i}}\right)\right]$

Jensen equality $< \ln\left(E\left[\frac{\Phi(x_i'\beta)^{y_i} (1 - \Phi(x_i'\beta))^{1-y_i}}{\Phi(x_i'\beta_0)^{y_i} (1 - \Phi(x_i'\beta_0))^{1-y_i}}\right]\right)$

LIE $= \ln\left(E\left[E\left[\frac{\Phi(x_i'\beta)^{y_i} (1 - \Phi(x_i'\beta))^{1-y_i}}{\Phi(x_i'\beta_0)^{y_i} (1 - \Phi(x_i'\beta_0))^{1-y_i}} \Big| x_i\right]\right]\right)$

$= \ln\left(E\left[\frac{\Phi(x_i'\beta)^{y_i}}{\Phi(x_i'\beta_0)^{y_i}} \underbrace{Pr(y_i = 1 | x_i)}_{\Phi(x_i'\beta_0)} + \frac{1 - \Phi(x_i'\beta)}{1 - \Phi(x_i'\beta_0)} \underbrace{Pr(y_i = 0 | x_i)}_{1 - \Phi(x_i'\beta_0)}\right]\right)$

$$l = \ln\left(E\left[\Phi(x_i'\beta) + 1 - \Phi(x_i'\beta)\right]\right)$$
$$= \ln E(1) = \ln 1 = 0$$
$$L(\beta) < L(\beta_0) \quad \forall \beta \neq \beta_0$$

$$E\left[(x_i'(\beta - \beta_0))^2\right] = (\beta - \beta_0)' E(x_i x_i')(\beta - \beta_0)$$

$$\hat\beta_{probit} = \underset{\beta}{\arg\max}\ \frac{1}{n} L_n(\beta)$$
$$\beta_0 = \underset{\beta}{\arg\max}\ L(\beta)$$

So if $\sup \left[\frac{1}{n} L_n(\beta) - L(\beta)\right] \xrightarrow{p} 0$ weak version: $\frac{1}{n} L_n(\beta) \xrightarrow{p} L(\beta)$

~~then~~ $\hat\beta_{probit} \xrightarrow{p} \beta$

Under Probit 0-1-2 (and additional regularity conditions)

a) $\hat\beta_{probit} \xrightarrow{p} \beta_0$

b) $\sqrt{n}(\hat\beta_{probit} - \beta_0) \xrightarrow{d} N(0, J^{-1})$

where $J = E\left[\frac{\partial}{\partial\beta} \ln f(y_i \mid x_i' \beta_0)\left\{\frac{\partial}{\partial\beta} \ln f(y_0 \mid x_1, \beta_0)\right\}'\right]$

information matrix $J = E\left[\frac{\phi(x_i'\beta_0)^2 x_i x_i'}{\Phi(x_i'\beta_0)(1 - \Phi(x_i'\beta_0))}\right]$

$$f(y_i \mid x_i, \beta) = \Phi(x_i'\beta)^{y_i}(1 - \Phi(x_i'\beta))^{1-y_i}$$
$$\ln f(y_i \mid x_i'\beta) = y_i \ln \Phi(x_i'\beta) + (1 - y_i)\ln(1 - \Phi(x_i'\beta))$$

$$\frac{\partial}{\partial\beta} \ln f(y_i \mid x_i, \beta) = y_i \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)} x_i + (1 - y_i)\frac{-\phi(x_i'\beta)}{1 - \Phi(x_i'\beta)} x_i$$
$$= \frac{y_i - \cancel{y_i\Phi(x_i'\beta)} - \Phi(x_i'\beta) + y_i\Phi(x_i'\beta)}{\Phi(x_i'\beta)(1 - \Phi(x_i'\beta))}$$

$$= \frac{y_i - \Phi(x_i'\beta)}{\Phi(x_i'\beta)(1 - \Phi(x_i'\beta))}\phi(x_i'\beta) x_i$$

$$J = E\left[\frac{(y_i - \Phi(x_i'\beta))^2}{\Phi(x_i'\beta_0)^2(1-\Phi(x_i'\beta_0))^2}\,\phi(x_i'\beta_0)^2 x_i x_i'\right]$$

$$E(\Phi(x_i'\beta_0)) \quad = E\left[\frac{y_i - 2y_i\Phi(x_i'\beta_0) + \Phi(x_i'\beta_0)^2}{\Phi(x_i'\beta_0)^2(1-\Phi(x_i'\beta_0))^2}\,\phi^2(x_i'\beta_0)\,x_i x_i'\right]$$

$$= E\left[\frac{\Phi(x_i'\beta_0) - \Phi(x_i'\beta_0)^2}{\Phi(x_i'\beta_0)^2(1-\Phi(x_i'\beta_0))^2}\,\phi(x_i'\beta_0)^2\,x_i x_i'\right]$$

$$= E\left[\frac{\phi(x_i'\beta_0)^2\,x_i x_i'}{\Phi(x_i'\beta_0)(1-\Phi(x_i'\beta_0))}\right]$$

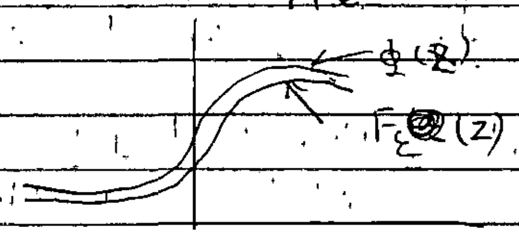
LOGIT

$$y_i^* = x_i'\beta + \varepsilon_i \qquad y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{if } y_i^* < 0 \end{cases}$$

$$\varepsilon_i | x_i \sim \text{logistic}$$

$$f_\varepsilon(\varepsilon|x) = \frac{e^{\varepsilon}}{(1+e^{\varepsilon})^2}$$

$$F_\varepsilon(\varepsilon|x) = \frac{e^{\varepsilon}}{1+e^{\varepsilon}}$$



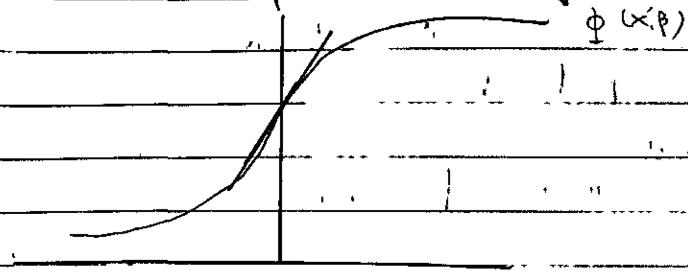
$$\Pr(y_i = 1|x_i) = \Pr(y_i^* \geq 0 \,|x_i)$$

$$= \Pr(\varepsilon_i \geq x_i'\beta_0 |x_i) = 1 - F_\varepsilon(-x_i'\beta_0 | x_i)$$

$$= 1 - \frac{e^{-x_i'\beta_0}}{1-e^{-x_i'\beta_0}} = \frac{1}{1+e^{-x_i'\beta_0}}$$

$$= \frac{e^{x_i'\beta_0}}{1+e^{x_i'\beta_0}}$$

$$J_{logit} = E\left[\frac{f_\varepsilon(x_i'\beta_0)^2 x_i x_i'}{F_\varepsilon(x_i'\beta_0)(1-F(x_i'\beta_0))}\right]$$



Semiparametric Binary choice

$y_i^* = x_i'\beta_0 + \varepsilon_i \quad y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{if } y_i^* < 0 \end{cases}$

$\varepsilon_i | x \sim N(0,1)$ (Probit)

$\sim$ logistic (Logit)

Assume med$(\varepsilon_i | x_i) = 0$

median

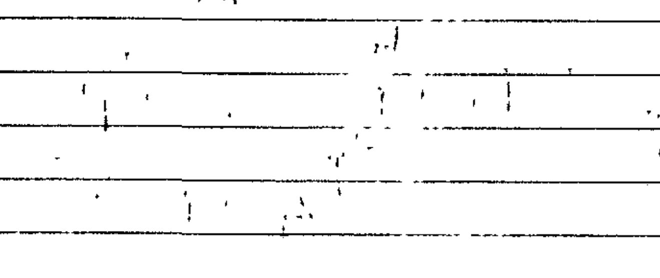
Suppose $x_i'\beta_0 > 0$.

$Pr(y_i = 1 | x_i) = Pr(x_i'\beta_0 + \varepsilon_i \geq 0 | x_i)$

$= Pr(\varepsilon_i \geq -x_i'\beta_0 | x_i)$

≥ 0.5

$\geq Pr(y_i = 0 | x_i)$



3.25th

Today

Binary choice

- Semi parametric Version
  Panel Version

Next

- Multinomial Choice

Manski (1985)

"Semiparametric Analysis of Discrete Response"
Journal of Econometrics, Vol. 27, p 313-333

$y = \mathbb{1}(y^* \geq 0)$ where $y^* = x'\beta_0 + \varepsilon$

Manski $\text{med}(\varepsilon | X) = 0$ $\Pr(\varepsilon \geq 0 | x) = 0.5$

If $x'\beta_0 \geq 0$, then

$$\Pr(y=1 | x) = \Pr(y^* \geq 0 | x)$$
$$= \Pr(x'\beta_0 + \varepsilon \geq 0 | x) = \Pr(\varepsilon \geq -x'\beta_0 | x)$$
$$\geq \tfrac{1}{2} \geq \Pr(y=0 | x)$$

Data $(y_1, x_1), \ldots, (y_n, x_n)$

Want to estimate β_0

Try to choose $\hat\beta$ so that

if $y_i = 1$, $x_i'\hat\beta > 0$

and if $y_i = 0$, $x_i'\hat\beta < 0$

choose β to solve

$$\max_\beta \frac{1}{n} \sum_{i=1}^n \left[y_i \mathbb{1}\{x_i'\beta > 0\} + (1-y_i) \mathbb{1}\{x_i'\beta < 0\} \right]$$

$\Longleftrightarrow$ equivalently

$$\max_\beta \frac{1}{n} \sum_{i=1}^n (2y_i - 1) \mathbb{1}\{x_i'\beta \geq 0\}$$

Maximum Score Estimation

Let $S(\beta) = E\left[(2y-1)\,\mathbb{1}\{x'\beta > 0\}\right]$

Exercise Show β_0 maximizes $S(\beta)$

Under conditions

(a) Consistency: $\hat{\beta}_{ms} \xrightarrow{P} \beta_0$

Advantages of Maximum Score

1) $\hat{\beta}_{ms}$ is consistent
2) Minimal restrictions on ε - distribution
3) Straightforward to extend to other quantiles than median.

Disadvantages of Maximum Score

1) Nonstandard Limit - Distribution AND slower than $\sqrt{n}$ -rate of covergence : $n^{\frac{1}{3}}(\hat{\beta}_{ms} - \hat{\beta}_0) \xrightarrow{d} \ldots$

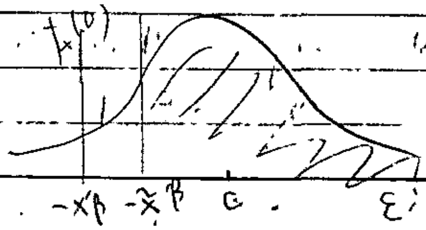
2) Interpretation : $\varepsilon | x$ can change with x

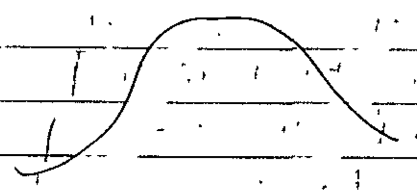
so $\frac{\partial P_r(y=1|x)}{\partial x}$ (marginal effects) and β Notes

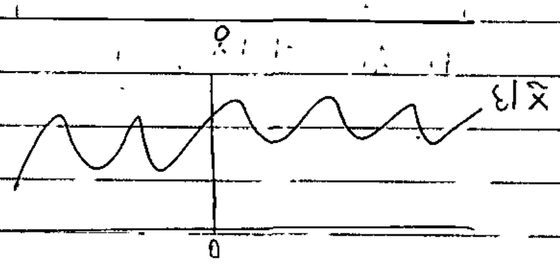
simply interpretted.

$P_r(y=1|x) = P_r[y^* \geq 0 \mid x] = P_r(\varepsilon \geq -x'\beta | x)$

$x \to \bar{x}$







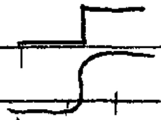
3) Cannot predict probabilities for x values outside the support of X.

ε indep x would take care of disadvantage (2)

Smoothed maximum score w/ Horowitz (1997)

$\mathbb{1}(z)$

$K(z)$



$$Pr(y=1 \mid x) = Pr(y^* \geq 0 \mid x)$$
$$= Pr(x'\beta_0 + \varepsilon \geq 0 \mid x)$$

Rivers - Vuong

Blundell - Smith

Blundell - Powell

Panel case

$$y_{it} = \begin{cases} 1 & \text{if } x_{it}'\beta_0 + \alpha_i + \varepsilon_{it} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Homogeneity / stationarity Assumption

$$\varepsilon_{it} \mid x_{is}, x_{it}, \alpha_i \sim \varepsilon_{is} \mid x_{is}, x_{it}, \alpha_i$$

Suppose $x_{is}'\beta_0 > x_{it}'\beta_0$

$$Pr(y_{it}=1 \mid x_{is}, x_{it}, \alpha_i)$$
$$= Pr(x_{it}'\beta_0 + \alpha_i + \varepsilon_{it} \geq 0 \mid x_{is}, x_{it}, \alpha_i)$$
$$= Pr(x_{it}'\beta_0 + \alpha_i + \varepsilon_{is} \geq 0 \mid x_{is}, x_{it}, \alpha_i)$$
$$\leq Pr(x_{is}'\beta_0 + \alpha_i + \varepsilon_{is} \geq 0 \mid x_{is}, x_{it}, \alpha_i)$$
$$= Pr(y_{is} = 1 \mid x_{is}, x_{it}, \alpha_i)$$

$$\geq \; = \; \leq$$

$$y_{it} = \mathbb{1}\{x_{it}'\beta_0 + \alpha_i + \varepsilon_{it} \geq 0\}$$
$$y_{is} = \mathbb{1}\{x_{is}'\beta_0 + \alpha_i + \varepsilon_{is} \geq 0\}$$

$$x_{it}'\beta_0 > x_{is}'\beta_0 \iff \Pr(y_{it}=1 \mid x_{is}, x_{it}, \alpha_i) > \Pr(y_{is}=1 \mid x_{is}, x_{it}, \alpha_i)$$

$$< \qquad\qquad <$$
$$= \qquad\qquad =$$

Estimate β via another "maximum score" criterion

$$\max_{\beta} \frac{1}{n} \sum_{i=1}^{n} \sum_{s=1}^{T-1} \sum_{t>s} (y_{is} - y_{it}) \, \mathrm{Sgn}(x_{is}'\beta - x_{it}'\beta)$$

Try to choose β so that $x_{it}'\hat{\beta} > x_{is}'\hat{\beta}$ if $y_{it} > y_{is}$

4.6th

Today
- Boot strap
- Parametric
- Non parametric

reading : Horowitz (2003)

Next:
- Multinomial Choice
- Generalized Method of Moments

reading : Nevo (2000)

Likelihood Framework

$Z_1, \dots, Z_n \overset{iid}{\sim} F_Z(z|\theta_0)$

e.g. $Z_1, \dots, Z_n \overset{iid}{\sim}$ Bernoulli (θ_0)

$Pr(z_i = 1 | \theta_0) = \theta_0$

$Pr(z_i = 0 | \theta_0) = 1 - \theta_0$

Probit..

e.g. $(y_1, x_1), \dots, (y_n, x_n)$ iid

$Pr(y_i | x_i) = \Phi(x_i' \beta_0)$

Usually we focus on estimators and confidence intervals

Estimator $T_n(Z_1, \dots, Z_n)$

$\hat{\beta}_{probit} = \arg\max_{\beta} \mathcal{L}(\beta | (y_1, x_1) \dots (y_n, x_n))$

T_n has a cdf $F_{T_n \theta_0}$

whatever transformation of random variables, create a new random variable (T_n) has new distribution.

Use F_{T_n, θ_0} to construct CIs

$\Rightarrow \sqrt{n}(T_n - \theta_0)$ is a simple transformation of T_n

$$0.95 = \Pr(a \le \sqrt{n}(T_n - \theta_0) \le b)$$

Find a, b, s.t.

$$= \Pr\left(\frac{a}{\sqrt{n}} \le T_n - \theta_0 \le \frac{b}{\sqrt{n}}\right)$$

$$= \Pr\left(-\frac{a}{\sqrt{n}} \ge \theta - T_n \ge -\frac{b}{\sqrt{n}}\right)$$

$$0.95 = \Pr\left(T_n - \frac{a}{\sqrt{n}} \ge \theta \ge T_n - \frac{b}{\sqrt{n}}\right)$$

$$= \Pr\left(T_n - \frac{b}{\sqrt{n}} \le \theta \le T_n - \frac{a}{\sqrt{n}}\right)$$

$\Rightarrow \left[T_n - \frac{b}{\sqrt{n}},\ T_n - \frac{a}{\sqrt{n}}\right]$ forms a 95% CI for θ_0

Problems:

1) T_n is often complicated transformation.

$\Rightarrow$ 2) $F_{T_n,\theta}$ is unknown.

3) θ_0 is unknown.

$$\sqrt{n}(\underset{\text{probit}}{\hat{\beta}} - \beta_0) \xrightarrow{d} \underbrace{N(0, J^{-1})}_{\text{cdf of } N(0, J^{-1})}$$

approxmites $F_{\sqrt{n}(T_n - \theta_0), \theta_0}$

$F_{\sqrt{n}(\hat{\beta} - \beta_0), \theta_0}$ probit

Simulation Approach to CI Construction

key steps

a) Choose θ_0

b) Draw $\tilde{Z}_1, \ldots, \tilde{Z}_n \overset{iid}{\sim} F_Z(z|\theta_0)$

c) Form $\tilde{T}_n = T_n(\tilde{Z}_1, \ldots, \tilde{Z}_n)$

d) Repeat b) and c)

$B = 10000$ times

$T_n^1, \ldots, T_n^B$

$\left[T_n - \frac{b}{\sqrt{n}},\ T_n - \frac{a}{\sqrt{n}}\right]$ forms 95% CI

Usually $T_n = \hat{\theta}_n$

→ Form histogram for $\sqrt{n}\,(T_n^{\,i} - \theta_0)$ $i = 1, \ldots, B$
→ Then use this histogram to approximate / obtain
a, b (Recall, $0.95 = \Pr(a \le \sqrt{n}\,(T_n - \theta_n) \le b)$)

Parametric Bootstrap
$Z_1, \ldots, Z_n$ iid P_θ $\| \; \prod_{i=1}^{n} f_Z(z|\theta)$
$\hat{\theta}_{MLE} = \arg\max_\theta \mathcal{L}(\theta | Z_1, \ldots, Z_n)$
a*) obtain $\hat{\theta}_{MLE}$ from sample
b*) Draw $Z_1^*, \ldots, Z_n^*$ iid $F_Z(z | \hat{\theta}_{MLE})$ ($P_{\hat{\theta}_{MLE}}$)
c*) Form $\hat{\theta}_{MLE}^* = \arg\max_\theta \mathcal{L}(\theta | Z_1^*, \ldots, Z_n^*)$
d*) Repeat b) & c) B times:
$\hat{\theta}_{MLE}^{*,1}, \ldots, \hat{\theta}_{MLE}^{*,B}$

Form histogram based on
$\sqrt{n}\,(\hat{\theta}_{MLE}^{*,1} - \hat{\theta}_{MLE})$,
$\vdots$
$\sqrt{n}\,(\hat{\theta}_{MLE}^{*,B} - \hat{\theta}_{MLE})$
→ use to approximate a, b
Often Bootstrap will be used to simulate
$$\frac{\sqrt{n}\,(\hat{\theta}^{*,1} - \hat{\theta}_{MLE})}{\hat{\sigma}_n^{*,1}}, \ldots, \frac{\sqrt{n}\,(\theta^{*,B} - \hat{\theta}_{MLE})}{\hat{\sigma}_n^{*,B}}$$

Find r, s such that
$$0.95 \le \Pr\left(r \le \frac{\sqrt{n}\,(T_n - \theta_0)}{\hat{\sigma}} \le s\right)$$
$$= \Pr\left(-r\frac{\hat{\sigma}}{\sqrt{n}} \ge \theta_0 - T_n \ge -s\cdot\frac{\hat{\sigma}}{\sqrt{n}}\right)$$
$$= \Pr\left(T_n - \frac{s\hat{\sigma}}{\sqrt{n}} \le \theta_0 \le T_n - \frac{r\hat{\sigma}}{\sqrt{n}}\right)$$
$\left[T_n - \frac{s\hat{\sigma}}{\sqrt{n}},\ T_n - \frac{r\hat{\sigma}}{\sqrt{n}}\right]$ forms 95% CI

4.8th

Today

- Bootstrap
- Intro Multinomial choice

Next

- Multinomial choice
- Nevo (2000)
- Greene | Multinomial Choice

Problem Set #4

$L(\beta)$ log likelihood for a single observation.

Recall $Z_1, \dots, Z_n \overset{iid}{\sim} F_Z(Z|\theta_0)$

Estimator: $T_n = T_n(Z_1, \dots, Z_n)$

95% CI $\left[T_n - \frac{a}{\sqrt{n}}, T_n - \frac{b}{\sqrt{n}}\right]$

where a, b are chosen to satisfy $0.95 = Pr(a \le \sqrt{n}(T_n - \theta_0) \le b)$

OR 95% CI $= \left[T_n - s \cdot \frac{\sigma}{\sqrt{n}}, T_n - r\frac{\sigma}{\sqrt{n}}\right]$

where r, s are chosen to satisfy

$.95 = Pr\left(r \le \frac{\sqrt{n}(T_n - \theta_0)}{\sigma} \le s\right)$

How do we obtain a, b or r, s?

1) Asymptotic Approx

$\sqrt{n}(T_n - \theta_0) \xrightarrow{d} N(0, \sigma^2) \Rightarrow a \approx -1.96\hat{\sigma}$, $b \approx 1.96\hat{\sigma}$

OR $\frac{\sqrt{n}(T_n - \theta_0)}{\sigma} \xrightarrow{d} N(0,1) \Rightarrow r = -1.96 \quad s \approx 1.96$

2) Bootstrap

Steps:

a) Obtain data $(Z_1, \dots, Z_n)$

Estimate parameter $\hat{\theta}_0 (\hat{\sigma})$

e.g Probit

Data : $(y_1, x_1), \ldots, (y_n, x_n)$

Estimates: $\hat{\beta}_{prob}$ $\hat{J}^{-1}$

b) Simulate a bootstrap sample.

Use random number generator to obtain bootstrap sample.

c) Parametric Bootstrap

$Z_1^*, \ldots, Z_n^* \underset{\sim}{iid} F_Z(z|\hat{\theta})$

d) Non parametric Bootstrap. Draw bootstrap with replacement from $(Z_1, \ldots, Z_n)$, OR

$Z_1^*, \ldots, Z_n^* \underset{\sim}{iid} \hat{F}_n$ (empirical cdf)

e.g Probit

(i) $\varepsilon_i^* \sim N(0,1)$, $y_i^* = \mathbb{1}\{x_i'\hat{\beta} + \varepsilon_i^* \geq 0\}$ for $i = 1, \ldots, n$

$\Rightarrow (y_1^*, x_1), \ldots, (y_n^*, x_n)$

(ii) Draw (y_i^*, x_i^*) from $(y_1, x_1) \ldots (y_n, x_n)$ (with replacement)

$F_n(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{Z_i \leq z\}$

(c) Use bootstrap sample $Z_1^*, \ldots, Z_n^*$ to form "bootstrap" estimates

$\hat{\theta}^*, \hat{\sigma}^{*2}$ e.g $\hat{\beta}^*_{prob}, \hat{J}^{*-1}$

(d) repeat steps b), c)

B times, where B is a large number

$\rightarrow \hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$

$\hat{\sigma}_1^{*2}, \ldots, \hat{\sigma}_B^{*2}$

Form

$$\mu_1^* = \sqrt{n}(\hat{\theta}_1^* - \hat{\theta}) \ \ldots \ \mu_B^* = \sqrt{n}.(\hat{\theta}_B^* - \hat{\theta})$$

OR

$$t_1^* = \sqrt{n}\frac{(\hat{\theta}_1^* - \hat{\theta})}{\hat{\sigma}_1^*}, \ \ldots \ t_B^* = \frac{\sqrt{n}(\hat{\theta}_B^* - \hat{\theta})}{\hat{\sigma}_B^*}$$

Re-order

$$\mu^*_{(1)} \le \mu^*_{(2)} \le \cdots \le \mu^*_{(B)}$$
$$t^*_{(1)} \le t^*_{(2)} \le \cdots \le t^*_{(B)}$$

Let $a^* = \mu^*([.025 B])$ $b^* = \mu^*([.975 B])$

OR $r^* = t^*_{([.025B])}$ $s^* = t^*_{([.975B])}$

e.g $B = 1000$ $a^* = \mu^*_{(25)}$ $b^* = \mu^*_{(975)}$

$r^* = t^*_{(25)}$ $s^* = t^*_{(975)}$

Finally 95% CI $\left[\hat{\theta} - \frac{b^*}{\sqrt{n}}, \hat{\theta} - \frac{a^*}{\sqrt{n}}\right]$

OR $\left[\hat{\theta} - s^*\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta} - r^*\frac{\hat{\sigma}}{\sqrt{n}}\right]$

Bootstrap Monte Carlo

Setup:

$X_1, \ldots, X_{20} \overset{iid}{\sim} Exp(5)$

Bootstrap Iteration : 1000

Non parametric Bootstrap: $X_1^*, \ldots, X_{20}^* \overset{iid}{\sim} F_n$

4. Booth

Today & Next

- Multi nominal Choice

Random Coefficient

Multi nominal logit

Generalized Method of Moments

Multinomial Choice

Outcome / Dependent variable

$y \in \{0, \dots, J\}$

e.g. - Mode of transportation
- Brand / size of cereal
- health insurance plan

Observe individuals $i = 1, \dots, n$

$(y_1, x_1) \dots (y_n, x_n)$

$$\Pr(y_i = j \mid x_i) = \frac{\exp(x_{ij}' \beta_0)}{\sum_{l=0}^{J} \exp(x_{il}' \beta_0)} \qquad j = 0, 1, \dots, J$$

where $x_i = (x_{i0}, x_{i1}, \dots, x_{iJ})$

Marginal Effects

$$\frac{\partial}{\partial x} \Pr(y_i = j \mid x_i) = \frac{\exp}{\sum_{l=0}^{J} \exp(x_{il}' \beta_0)} - \frac{[\exp(x_{ij}' \beta_0)]^2 \beta_{0,k}}{[\sum_{l=0}^{J} \exp(x_{il}' \beta_0)]^2}$$

$$= \frac{\exp(x_{ij}' \beta_0)}{\sum_{l=0}^{J} \exp(x_{il}' \beta_0)} \left[1 - \frac{\exp(x_{ij}' \beta_0)}{\sum_{l=0}^{J} \exp(x_{il}' \beta_0)}\right] \beta_k$$

$$= \Pr(y_i = j \mid x_i)[1 - \Pr(y_i = j \mid x_i)] \beta_k$$

$$\frac{\partial}{\partial x_{i,l,k}} \Pr(y = j \mid x_i) \; (j \neq l) = \frac{\exp(x_{ij}' \beta_0) \exp(x_{il}' \beta_0) \beta_l}{-\left[\sum_{l=0}^{J} \exp(x_{il}' \beta_0)\right]^2}$$

$$= -\Pr(y_i = j \mid x_i) \Pr(y_i = l \mid x_i) \beta_l$$

- log odds Ratio

$$\ln\left(\frac{\Pr(y_i = j \mid x_i)}{\Pr(y_i = l \mid x_i)}\right) = \ln\left(\frac{\exp(x_{ij}' \beta_0)}{\exp(x_{il}' \beta_0)}\right)$$

$$= (x_{ij} - x_{il})' \beta_0$$

$$\Pr(y_i = j \mid y_i = j \text{ or } y_i = l, x_i)$$

$$= \frac{\Pr(y_i = j \mid x_i)}{\Pr(y_i = j \mid x_i) + \Pr(y_i = l \mid x_i)} = \frac{\dfrac{\exp(x_{ij}'\beta_0)}{\sum_{l=0}^{J} \exp(x_{il}'\beta_0)}}{\dfrac{\exp(x_{ij}'\beta_0)}{\sum_{l=0}^{J} \exp(x_{ij}'\beta_0)} + \dfrac{\exp(x_{il}'\beta_0)}{\sum_{j=0}^{J} \exp(x_{il}'\beta_0)}}$$

$$= \frac{\exp(x_{ij}'\beta_0)}{\exp(x_{ij}'\beta_0) + \exp(x_{il}'\beta_0)} = \frac{\exp((x_{ij} - x_{il})'\beta_0)}{1 + \exp((x_{ij} - x_{il})'\beta_0)}$$

Conditional logit

Random Utility:

$$y_{ij}^* = x_{ij}'\beta + \varepsilon_{ij}, \quad j = 0, \dots J$$

x_{ij} = characteristics of choice j for consumer

$$y_i = \arg\max_j y_{ij}^*$$

$\varepsilon_{i0}, \dots \varepsilon_{iJ}$ jointly indep

$\varepsilon_{ij} \mid x_i \sim$ extreme value Type I

$$\Pr(y_i = j \mid x_i) = \frac{\exp(x_{ij}'\beta_0)}{\sum_{l=0}^{J} \exp(x_{il}'\beta_0)}$$

$$= \frac{\exp(x_{ij}'\beta_0)}{\exp(x_{i0}'\beta_0) + \dots + \exp(x_{iJ}'\beta_0)}$$

$$\exp(x_{ij1}\beta_{01} + x_{ij2}\beta_{02}) = \exp(x_{ij1}\beta_{01})\exp(x_{ij2}\beta_{02})$$

$$x_{i0} = \begin{pmatrix} inc_i \text{ (income)} \\ - \\ \vdots \\ x_{i02} \end{pmatrix} \quad x_{i1} = \begin{pmatrix} 0 \\ inc_i \\ 0 \\ \vdots \\ x_{i12} \end{pmatrix} \quad \dots x_{iJ} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ inc_i \\ x_{iJ2} \end{pmatrix}$$

$$\delta_0 = \begin{pmatrix} \delta_{00} \\ \vdots \\ \delta_{0J} \\ \beta_{02} \end{pmatrix}$$

$$X_{i0}'\delta_0 = \text{inc}_i\, \delta_{00} + X_{i20}\,\beta_{02}$$
$$X_{i1}'\delta_0 = \text{inc}_i\, \delta_{01} + X_{i21}\,\beta_{02}$$
$$\vdots$$
$$X_{iJ}'\delta_0 = \text{inc}_i\, \delta_{0J} + X_{i2J}\,\beta_{02}$$

then above

$$y_{ij}^* = x_{ij}'\beta_0 + W_i'\delta_{0j} + \varepsilon_i$$
$$j = 0, 1, \ldots, J$$

April 15th

TODAY
Multinomial choice
BLP
(Nevo 2000)
GMM Intro

NEXT
GMM
reading: Greene GMM chapter.
Newey McFadden

Berry, Levinsohn, Pakes (BLP)
Nevo (2000)

(i) Consumers
(j) Products
(t) Markets

Indirect Utility

marginal utility of income for i

individual taste coeffecient

Utility

$$U_{ijt} = (y_i - p_{jt})\alpha_i + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt} \leftarrow \text{Mean zero other unobserved utility determinants}$$

income for i — Price of good j in market t — product chars — unobsd product char

Individual Preference Model:

$$\begin{pmatrix}\alpha_i\\ \beta_i\end{pmatrix} = \begin{pmatrix}\alpha\\ \beta\end{pmatrix} + \Pi D_i + \Sigma v_i$$

observed demographic variables — unobserved chars

Assume $v_i \sim N(0, I)$

$\Sigma v_i \sim N(0, \Sigma)$

y_i indep D_i

Outside good / option

$$U_{i0t} = \alpha_i y_i + \xi_{0i} + \Pi_0 D_i + \sigma_0 v_{i0} + \varepsilon_{i0t}$$

$$U_{ijt} = (y_i - p_{jt})\alpha_i + x_{jt}\beta_i + \xi_{jt} + \cancel{} + \varepsilon_{ijt}$$

$$= \alpha_i y_i + \begin{pmatrix}-p_{jt}\\ x_{jt}\end{pmatrix}' \begin{pmatrix}\alpha_i\\ \beta_i\end{pmatrix} + \xi_{jt} + \varepsilon_{ijt}$$

$$= \alpha_i y_i + \underbrace{\left[\begin{pmatrix}-p_{jt}\\ x_{jt}\end{pmatrix}' \begin{pmatrix}\alpha\\ \beta\end{pmatrix} + \xi_{jt}\right]}_{=\delta_{jt}\ (\text{mean utility})} + \underbrace{\begin{pmatrix}-p_{jt}\\ x_{jt}\end{pmatrix}' [\Pi D_i + \Sigma v_i]}_{\mu_{ijt}} + \varepsilon_{ijt}$$

Instruments Z_{jt}

Assume $E[Z_{jt}\ \xi_{jt}]=0$

ASIDE

OLS / IV WORLD

$y_i = x_i'\beta + \varepsilon_i$

OLS: $0 = E(x_i \varepsilon_i)$

IV: $0 = E(z_i \varepsilon_i)$

$\therefore$ choose $\hat\beta$ $= \hat{E}[z_i (y_i - x_i'\beta_0)]$

s.t. $0 = \frac{1}{n}\sum_{i=1}^{n} z_i (y_i - x_i'\beta_0)$

$E[Z_{jt}\ \xi_{jt}\ (y_i, x_{jt}, P_i, \alpha_i \beta, \pi, \Sigma)]$

choose $\hat\alpha\ \hat\beta, \ldots$ s.t.

$0 = \frac{1}{nJ}\sum_{jt} Z_{jt}\ \xi_{jt}\ (\ldots \hat\alpha, \hat\beta, \ldots)$

April 20th

- TODAY
- BLP Empirical Results
- GMM
- Next

Quantile Regression

$0 = E[\xi_{jt}, \zeta_{jt}]$

Generalized Method of Moments

Examples

A) Regression: $y_i = x_i'\beta + \varepsilon_i$

(OLS 2) or (OLS 2') $\Rightarrow 0 = E(x_i \varepsilon_i) = E(x_i(y_i - x_i'\beta))$

B) Instrumental Variables

(IV2) or (IV2')

$\Rightarrow 0 = E(z_i \varepsilon_i) = E(z_i(y_i - x_i'\beta))$

C) Panel Data

$y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it}$

$-(\bar{y}_i = \bar{x}_i'\beta + \alpha_i + \bar{\varepsilon}_i)$

$\underbrace{y_{it} - \bar{y}_i}_{\tilde{y}_{it}} = \underbrace{(x_{it} - \bar{x}_i)'}_{\tilde{x}_{it}}\beta + \underbrace{\varepsilon_{it} - \bar{\varepsilon}_i}_{\tilde{\varepsilon}_{it}}$

$\Rightarrow 0 = E(\tilde{x}_{it}\tilde{\varepsilon}_{it})$

$= E(\tilde{x}_{it}(\tilde{y}_{it} - \tilde{x}_{it}'\beta))$

D) Nonlinear Regression

$E(y_i|x_i) = \Phi(x_i'\beta)$

$(y_i = \Phi(x_i'\beta) + \varepsilon_i)$

(Orthogonality $E(\varepsilon_i|x_i) = 0$.)

$\Rightarrow 0 = E[h(x_i)\varepsilon_i]$

$= E[h(x)(y_i - \Phi(x_i'\beta))]$

where $h(\cdot)$ is any function of x

e.g. $h(x_i) = \frac{\partial}{\partial\beta}\Phi(x_i'\beta)$

E) MLE

$L(\theta) = \ln f(z|\theta)$

$E[L(\theta)] = E[\ln f(z|\theta)]$

θ maximizes $E[L(\theta)]$

FOC: $0 = E\left[\frac{\partial}{\partial\theta}L(\theta_0)\right] = E\left[\frac{\partial}{\partial\theta}\ln f(z|\theta_0)\right]$

F) BLP $0 = E[\xi_{jt} z_{jt}]$

GMM "Model"

$0 = E[g(z_i, \theta_0)]$

Observe $z_1, z_2, \ldots$ iid

$g()$ known function θ_0 unknown

$Z_i = (y_i, x_i)$

Ex $g(z_i, \theta) = X_i (y_i - x_i'\theta)$

OR

Ex. $g(z_i, \theta) = h(x_i)'(y_i - \Phi(x_i'\theta))$

Idea: observe Data $Z_1, \dots, Z_n$
want to estimate θ_0. Find $\hat{\theta}_{GMM}$ such that
$$0 \approx \frac{1}{n}\sum_{i=1}^{n} g(Z_i, \hat{\theta}_{GMM})$$

g is $L \times 1$
θ is $k \times 1$

$L = k$ (just id) identity

Often will be able to find $\hat{\theta}_{GMM}$ to solve
$0 = \frac{1}{n}\sum_{i=1}^{n} g(Z_i, \hat{\theta}_{GMM})$ exactly
(Method of Moments)

$L > k$ (over identified)
$$\min_{\theta} \left[\frac{1}{n}\sum_{i=1}^{n} g(Z_i, \theta)\right]' \left[\frac{1}{n}\sum_{i=1}^{n} g(Z_i, \theta)\right]$$

GMM

Suppose (i) $W_n \xrightarrow{p} W$ p.s.d $\theta_0 \in \text{int}(\Theta)$, Θ compact

(ii) Regularity conditions on g: $g(Z, \theta)$ continuously differentiable
$\| E[\|g(Z, \theta_0)\|] < \infty$ $E\left[\sup_{\theta \in \Theta} \|g(Z, \theta)\|\right] < \infty$
$E\left[\sup_{\Theta} \|\frac{\partial}{\partial \theta} g(Z, \theta_0)\|\right] < \infty$

(GMM 0) $Z_1, Z_2,$ iid

(GMM 1) $G'WG$ non singular, where $G = E\left[\frac{\partial}{\partial \theta} g(Z, \theta)\right]$

(GMM 2) $E[g(Z, \theta_0)] = 0$ (and $WE[g(Z, \theta)] = 0$, iff $\theta = \theta_0$)

(GMM 3) $\Omega = E[g(Z, \theta_0) g(Z, \theta_0)']$ finite

Then $\hat{\theta}_{GMM}$ solves
W_n weight matrix

$$\min_\theta \left[\frac{1}{n}\sum_i g(Z_i,\theta)\right]' W_n \left[\frac{1}{n}\sum_i g(Z_i,\theta)\right]$$

(a) $\hat{\theta}_{GMM} \xrightarrow{P} \theta_0$

(b) $\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(0, (G'WG)^{-1} G'W\Omega WG (G'WG)^{-1})$

$j = 1, \dots, q$ players
π : pay off function
a : actions (continuous)
x : other determinants of payoffs
J : information set of i

$$0 = E\Big[\underbrace{\frac{\partial}{\partial a_i}\pi(a_i, a_{-i}, x_i, \theta_0)}_{g(Z,\theta_0)} \Big| J_i\Big] = 0$$

Weight Matrix W_n

Example OLS / Regression

$$y_i = x_i'\beta + \varepsilon_i$$

$$0 = E(x_i\varepsilon_i) = E[x_i(y_i - x_i'\beta_0)]$$

OLS: $\hat{\beta}_{OLS}$ minimizes $\sum_{i=1}^n (y_i - x_i'\beta)^2$

$$\Rightarrow \hat{\beta}_{OLS} = \left(\frac{1}{n}\sum x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n x_i y_i$$

GMM

choose $\hat{\beta}_{GMM.OLS}$ to solve

$$\min_\beta \Big[\underbrace{\frac{1}{n}\sum_i x_i(y_i - x_i'\beta)}\Big]' W_n \left[\frac{1}{n}\sum_i x_i(y_i - x_i'\beta)\right]$$

$$-2\beta\left[\frac{1}{n}\sum_i x_i x_i'\right]' W_n \left[\frac{1}{n}\sum_i x_i y_i\right] +$$

$$\beta'\left[\frac{1}{n}\sum_i x_i x_i'\right]' W_n \left[\frac{1}{n}\sum_i x_i x_i'\right]\beta$$

FOC :

$$0 = -2\left[\frac{1}{n}\sum x_i x_i'\right]' W_n \left[\frac{1}{n}\sum_i x_i y_i\right] + 2\left[\frac{1}{n}\sum_i x_i x_i'\right]' W_n \left[\frac{1}{n}\sum_i x_i x_i'\right]\hat\beta$$

$$0 = -2A + 2C\hat\beta$$

$$\hat\beta = C^{-1}A$$

$$\hat\beta_{GMM-OLS} = \left[\left(\frac{1}{n}\sum x_i x_i'\right)' W_n \left(\frac{1}{n}\sum x_i x_i'\right)\right]^{-1} \left(\frac{1}{n}\sum x_i x_i'\right) W_n \frac{1}{n}\sum x_i y_i$$

$$= \left(\frac{1}{n}\sum_i x_i x_i'\right)^{-1} W_n^{-1} \left(\frac{1}{n}\sum x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum x_i x_i'\right) W_n \frac{1}{n}\sum x_i y_i$$

$$= \left(\frac{1}{n}\sum_{i=1} x_i x_i\right)^{-1} \frac{1}{n}\sum x_i y_i = \hat\beta_{OLS}$$

Example IV

$$y_i = x_i'\beta + \varepsilon_i \qquad \hat\beta_{2SLS}$$

$$0 = E(z_i \varepsilon_i)$$

$$\hat\beta_{2SLS} = \left[\left(\frac{1}{n}\sum_i x_i z_i'\right)\left(\frac{1}{n}\sum z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_i z_i x_i\right)\right]^{-1} \left(\frac{1}{n}\sum_i x_i z_i'\right)\left(\frac{1}{n}\sum z_i z_i'\right)^{-1}\frac{1}{n}\sum z_i y_i$$

$$(a-b)' W_n (a-b)$$

GMM : choose $\hat\beta_{GMM-IV}$ to solve

$$\min_\beta \left[\frac{1}{n}\sum_i z_i(y_i - x_i'\beta)\right]' W_n \left[\frac{1}{n}\sum_i z_i(y_i - x_i'\beta)\right]$$

$$\Leftrightarrow \min_\beta \left[\frac{1}{n}\sum z_i y_i - \frac{1}{n}\sum z_i x_i'\beta\right]' W_n \left[\frac{1}{n}\sum_i z_i y_i - \frac{1}{n}\sum z_i x_i'\beta\right]$$

FOC $0 = -2\left[\frac{1}{n}\sum z_i x_i'\right]' W_n \left[\frac{1}{n}\sum_i z_i y_i\right] + 2\hat\beta'\left[\frac{1}{n}\sum_i z_i x_i'\right]' W_n \left[\frac{1}{n}\sum_i z_i x_i'\right]$

$$\hat\beta_{GMM,IV} = \left[\left(\frac{1}{n}\sum_{i=1}(x_i z_i')\right) W_n \left(\frac{1}{n}\sum_{i=1} z_i x_i'\right)\right]^{-1} \left(\frac{1}{n}\sum_{i=1} x_i z_i'\right) W_n \frac{1}{n}\sum_{i=1} z_i y_i$$

2SLS $= W_n = \left(\frac{1}{n}\sum z_i z_i'\right)^{-1}$

Optimal Weight Matrix Choice

$$\sqrt{n}\,(\hat\theta_{GMM} - \theta) \xrightarrow{d} N\big(0, \underbrace{(G'WG)^{-1} G'W\Omega W G (G'WG)^{-1}}_{AVAR(GMM)}\big)$$

$$W = \Omega^{-1} \Rightarrow AVAR = (G'\Omega^{-1}G)^{-1}$$

$$\Omega = E[g(z_i,\theta_0)\, g(z_i,\theta_0)']$$

$$(IV) \quad = E[z_i \varepsilon_i (z_i \varepsilon_i)'] \qquad \underset{\text{under homosked}}{= \sigma^2}$$

$$= E[\varepsilon_i^2 z_i z_i'] = [E(\varepsilon_i^2 | z_i) z_i z_i']$$

Homosked $E(\varepsilon_i^2 | z_i) = \sigma^2$ $\quad = \sigma^2 E[z_i z_i']$

$$W = \Omega^{-1} = \frac{1}{\sigma^2} E(z_i z_i')^{-1}$$

IV

$$\Omega = E(\varepsilon_i^2 z_i z_i')$$

Step 1.

use $\beta_{2SLS} \rightarrow \hat{\varepsilon} = y_i - x_i'\hat{\beta}_{2SLS}$

$\rightarrow W_n = (\frac{1}{n}\sum \hat{\varepsilon}_i^2 z_i z_i')^{-1}$

Step 2: Do GMM using W_n?

April 27th.

Today & Next

Quantile Regression

Koenker - Hallock (01)

Arios, Hallock, Sosa (01)

GMM

· Moment " or " generalized residual" function $g(z,\theta)$

$$0 = E[g(z,\theta)]$$

One step: Choose weight matrix W

Find $\hat{\theta}_{GMM\text{-}one}$ to solve $\min_\theta [\frac{1}{n}\sum_{i=1}^{n} g(z,\theta)]' W [\frac{1}{n}\sum_{i=1}^{n} g(z,\theta)]$

Two step: Use $\hat{\theta}_{GMM\text{-}one}$ to estimate optimal weight matrix

$$\hat{W}_{opt} = \left[\frac{1}{n} \sum g(Z_i, \hat\theta_{GMM\,one}) g(Z_i, \hat\theta_{GMM\,one})'\right]^{-1}$$

(approximates $\Omega^{-1} = E[g(Z, \theta_0) g(Z, \theta_0)']^{-1}$)

Find $\hat\theta_{GMM\text{-}opt}$ to solve $\min_\theta \left[\frac{1}{n}\sum_i g(Z_i, \theta)\right]' \hat{W}_{opt} \left[\frac{1}{n}\sum_i g(Z_i, \theta)\right]$

$\hat\theta_{GMM\text{-}one} \xrightarrow{p} \theta_0$, $\hat\theta_{GMM\text{-}opt} \xrightarrow{p} \theta_0$

$\sqrt{n}(\hat\theta_{GMM\text{-}one} - \theta_0) \xrightarrow{d} N(0, (G'WG)^{-1} G' W \Omega W G (G'WG)^{-1})$

$\sqrt{n}(\hat\theta_{GMM\text{-}opt} - \theta_0) \xrightarrow{d} N(0, (G'\Omega^{-1}G)^{-1})$

Recall: $G = E\left[\frac{\partial}{\partial\theta} g(Z, \theta_0)\right]$

Quantile Regression

$Z_1, Z_2, \ldots$ Z iid, q_τ

$= P_r(Z_i \le q_\tau) = \tau$

$q_{0.5} = P_r(Z_i \le q_{.5}) = \frac{1}{2}$

1.0, 1.2, 1.8, 2.2, 2.7, 3.2, 3.3, 3.5, 4.1

$\hat q_{.25}$ $\uparrow$ $\hat q_{.5}$

M#1 Find $\hat q_{0.5}$ to solve

$$0 = \sum_{i=1}^{n} m(Z_i, q) \quad \text{where } m(Z_i, \theta) = \mathbb{1}\{Z \le \theta\} - \mathbb{1}\{Z > \theta\}$$

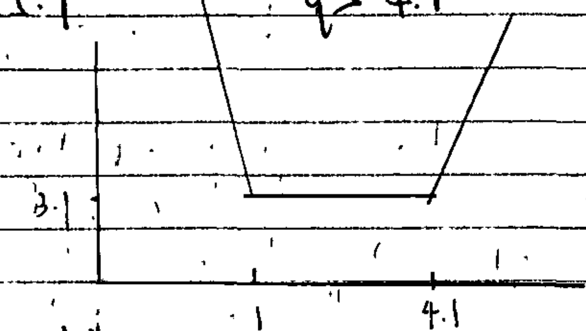
M#2 Find $\hat q_{0.5}$ to solve $\min_q \sum_{i=1}^{n} |Z_i - q|$

e.g. Two data points 1.0, 4.1

$$|Z_1 - q| + |Z_2 - q| = |1 - q| + |4.1 - q|$$

$$= \begin{cases} (1-q) + (4.1-q) & q < 1.0 \\ q - 1 + 4.1 - q & 1 \le q \le 4.1 \\ (q-1) + (q-4.1) & q > 4.1 \end{cases}$$

$$= \begin{cases} 5.1 - 2q & q < 1 \\ 3.1 & 1 \le q \le 4.1 \\ 2q - 5.1 & q > 4.1 \end{cases}$$



e.g. Three data pts: 1, 2.7, 4.1

$$\sum_{i=1}^{3} |z_i - q|$$

$$= |1-q| + |4.1-q| + |2.7-q|$$

~~minimum~~ minimizer $\Rightarrow \hat{q}_{.5} = 2.7$

Estimate q_τ:

(M#1) Find $\hat{q}_\tau$ to solve $0 = \sum_{i=1}^{n} m_\tau(Z_i, q)$

where $m(Z, \theta) = [(1-\tau)\mathbb{1}\{Z \le \theta\} - \tau \mathbb{1}\{Z > \theta\}]$

(M#2): Find $\hat{q}_\tau$ to solve

$$\min_q \sum_{i: Z_i \le q} |Z_i - q|(1-\tau) + \sum_{i: Z_i > q} |Z_i + q|\tau$$

Quantile Regression

$$q_\tau(y|x) = x'\beta_\tau$$

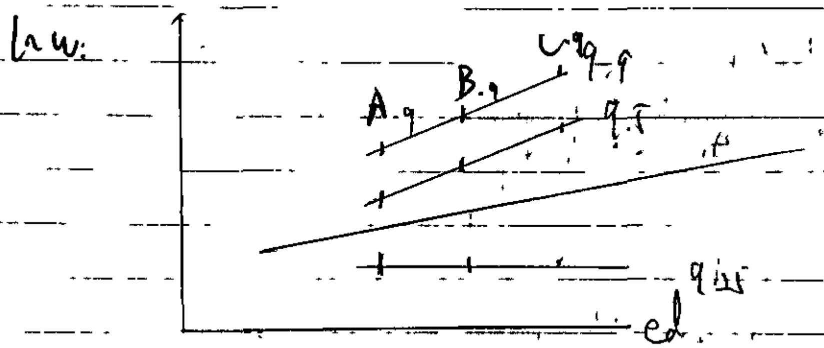
$$\Pr(y \le x'\beta_\tau | x) = \tau$$

$$y = x'\beta_\tau + \varepsilon \quad \text{where } \Pr(\varepsilon \le 0 | x) = \tau$$

$$q_\tau(\varepsilon|x) = 0$$

Example $\ln w_i = \beta_0 + \beta_1 ed_i + \beta_2 exp_i + \varepsilon_i$

Fix $exp_i = 5$



Pt A.q is
q.q(ln w| ed = 12, exp = 5)

Pt C.q is
q.q(ln w| ed = 20, exp = 5)

slope of line connecting A.q and C.q is $\Delta\beta_{1,\tau=.q}$

For $\tau = 1/2$.
Find $\hat{\beta}_{\tau=\frac{1}{2}}$ to solve $\min_{\beta} \sum_{i=1}^{\wedge} |y_i - x_i'\beta|$

For general τ,
Find $\hat{\beta}_\tau$ to solve $\min_{\beta} \sum_{i=1}^{\wedge} [(1-\tau)|y_i - x_i'\beta| \mathbb{1}\{y_i \le x_i'\beta\} + \tau |y_i - x_i'\beta| \mathbb{1}\{y_i > x_i'\beta\}]$

Schedule
Mon May 4 PS due
Wed May 6 Review (in class)
Mon May 11 Final Exam
2:45 – 4:45 pm
Soc. Sci #5206

TODAY

- Quantile Regression & GMM
- Ariacietal paper

NEXT

Forecast Overview

Quantile Regression.

$$q_\tau(y|x) = x'\beta_\tau \iff \Pr(y \le x'\beta_\tau|x) = \tau$$

$$\iff q_\tau(\varepsilon_\tau|x) = 0 \iff \Pr(\varepsilon_\tau \le 0|x) = \tau$$

where $\varepsilon_\tau = y - x'\beta_\tau$

QR Let $\hat\beta_\tau$ solves

$$\min_\beta \sum_{i=1}^n (1-\tau)|y_i - x_i'\beta| \mathbb{1}\{y_i \le x_i'\beta\} + \tau|y_i - x_i'\beta| \mathbb{1}\{y_i > x_i'\beta\}$$

$$\rho_\tau(\varepsilon) = (1-\tau)|\varepsilon| \mathbb{1}\{\varepsilon \le 0\} + \tau|\varepsilon| \mathbb{1}\{\varepsilon > 0\}$$

$$= \tau\varepsilon \mathbb{1}\{\varepsilon > 0\} - (1-\tau)\varepsilon\mathbb{1}\{\varepsilon \le 0\}$$

$$\Rightarrow \min_\beta \sum_{i=1}^n \rho_\tau(y_i - x_i'\beta)$$

Let $m(y_i, x_i\beta) = [(1-\tau)\mathbb{1}\{y_i - x_i\beta\} - \tau\mathbb{1}\{y_i > x_i\beta\}]x_i$

let $\hat\beta_\tau$ solve

$$0 = \frac{1}{n}\sum_i m(y_i, x_i, \hat\beta_\tau)$$

equivalently

$$\min_{\beta_\tau} \frac{1}{n}\sum_{i=1}^n m(y_i, x_i, \beta_\tau)' W \left[\frac{1}{n}\sum_{i=1}^n m(y_i, x_i, \beta_\tau)\right]$$

$$\sqrt{n}(\hat\beta_\tau - \beta_\tau) \xrightarrow{d} N(0, (G'G)^{-1} G'\Omega G (G'G)^{-1})$$

where $\Omega = E[m(y_i, x_i, \beta_\tau) m(y_i, x_i, \beta_\tau)']$

and $G = E\left[\frac{\partial}{\partial\beta_\tau} m(y_i x_i, \beta_\tau)\right]$

$$G = \frac{\partial}{\partial\beta} E[m(y_i, x_i, \beta)]\Big|_{\beta=\beta_0}$$

$$\Omega = E[((1-\tau)^2 \mathbb{1}\{y_i \le x_i'\beta_\tau\} + \tau^2 \mathbb{1}\{y_i > x_i'\beta_\tau\}) x_i x_i']$$

$$= E[((1-\tau)^2 \Pr(y_i \le x_i'\beta_\tau | x_i) + \tau^2 \Pr(y_i > x_i'\beta_\tau | x_i)) x_i x_i']$$

$$\Omega = E[\underbrace{((1-\tau)^2 \tau + \tau^2 (1-\tau))}_{=\tau(1-\tau)} x_i x_i'] = \Omega$$

$$= \tau(1-\tau) E[x_i x_i']$$

$1-\tau$

$$E[m(y_i, x_i, \beta)]$$
$$= E[((1-\tau)\mathbb{1}\{y_i \le x_i'\beta\} - \tau \mathbb{1}\{y_i > x_i'\beta\}) x_i]$$
$$= E[((1-\tau)\mathbb{1}\{\varepsilon_i \le x_i'(\beta-\beta_\tau)\} - \tau \mathbb{1}\{\varepsilon_i > x_i'(\beta-\beta_\tau)\}) x_i]$$
$$= E[((1-\tau)\Pr(\varepsilon_i \le x_i'(\beta-\beta_\tau) | x_i) - \tau \Pr(\varepsilon > x_i'(\beta-\beta_\tau) | x_i)) x_i]$$

Let $m(y_i, x_i, \beta) = [(1-\tau)\mathbb{1}\{y_i \le x_i'\beta\} - \tau\mathbb{1}\{y_i > x_i'\beta\}] x_i$

$$E[m(y_i, x_i, \beta)] = E[((1-\tau) F_{\varepsilon_i|x_i}(x_i'(\beta-\beta_\tau)) - \tau(1 - F_{\varepsilon_i|x_i}(x_i'(\beta-\beta_\tau)))) x_i]$$

$$\frac{\partial}{\partial\beta} E[m(y_i, x_i, \beta)] = E[\{(1-\tau) f_{\varepsilon_i|x_i}(x_i'(\beta-\beta_\tau)) + \tau f_{\varepsilon_i|x_i}(x_i'(\beta-\beta_\tau))\} x_i x_i']$$

$$G = \frac{\partial}{\partial\beta} E[m(y_i, x_i, \beta)]\Big|_{\beta=\beta_\tau} = E[\{(1-\tau) f_{\varepsilon_i|x_i}(0) + \tau f_{\varepsilon_i|x_i}(0)\} x_i x_i']$$

$$G = E[f_{\varepsilon_i|x_i}(0) x_i x_i']$$

when G is symmetric and positive definite

$$(G'G)^{-1} G' \Omega G (G'G)^{-1} = G^{-1} \Omega G^{-1}$$

$$\sqrt{n}(\hat\beta_\tau - \beta_\tau) \xrightarrow{d} N(0, \tau(1-\tau) E[f_{\varepsilon_i|x_i}(0) x_i x_i']^{-1} E(x_i x_i') E[f_{\varepsilon_i|x_i}(0) x_i x_i']^{-1})$$

$$D_i = \begin{cases} 1 & \text{if take new drug} \\ 0 & \text{if take old drug} \end{cases}$$

Y denotes cholesterol

Potential $\begin{cases} Y_{1,i} = \text{cholesterol if given new drug } (D_i = 1) \\ Y_{0,i} = \text{cholesterol if given old drug } (D_i = 0) \end{cases}$
Outcome

$Y_{1,i} - Y_{0,i}$

Average Treatment Effect

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

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

classify into group with random method

$Y_{1,i} - Y_{0,i}$

Quantile Treatment Effect $= q_\tau(Y_{1,i} - Y_{0,i})$

$\neq q_\tau(Y_{1,i}) - q_\tau(Y_{0,i})$ depend on joint distribution

5.4

TODAY

Quantile Regression

Interpretation

- Accident paper

Next

Review

FINAL

cholesterol Drug

$Y_{1,i}$ = potential cholesterol if receive new drug

$Y_{0,i}$ = old drug

$Y_{1,i} - Y_{0,i}$ $E(Y_{1,i})$

$E(Y_{1,i}) - E(Y_{0,i})$

$q_\tau(Y_{1,i} - Y_{0,i})$

$q_\tau(Y_{1,i})$ $q_\tau(Y_{0,i})$

$q_\tau(Y_{1,i} - Y_{0,i}) \neq q_\tau(Y_{1,i}) - q_\tau(Y_{0,i})$

	Pop #1			Pop #2		
q	Y_0	Y_1	$Y_1 - Y_0$	Y_0	Y_1	$Y_1 - Y_0$
90	170	150	-20	170	110	-60
70	160	140	-20	160	120	-40
50	150	130	-20	150	130	-20
30	140	120	-20	140	150	10
10	130	110	-20	130	140	10

$E(Y_0) = 150$ $E(Y_1) = 130$ $E(Y_1 - Y_0) = -20$

$E(Y_1) - E(Y_0) = -20$

Pop#1 $q_{.3}(Y_1) = 140$ $q_{.3}(Y_0) = 120$ $q_{.3}(Y_1 - Y_0) = -20$

$q_{.3}(Y_1) - q_{.3}(Y_0) = q_{.3}(Y_1 - Y_0)$

Pop#2 $E(Y_0) = 150$ $E(Y_1) = 130$

$q_{.3}(Y_0) = 140$ $q_{.3}(Y_1) = 120$

$E(Y_1 - Y_0) = -20$

$q_{.3}(Y_1 - Y_0) = -40$

$$Y_i = \beta_{0,.3} + \beta_{1,.3} T_i + \varepsilon_{i;.3}$$

$$T_i = \begin{cases} 1 & \text{if new drug} \\ 0 & \text{if old drug} \end{cases}$$

$$q_{.3}(Y_i \mid T_i = 0) = \beta_{0,.3}$$

$$q_{.3}(Y_i \mid T_i = 1) = \beta_{0,.3} + \beta_{1,.3}$$

$$\beta_{1,.3} = q_{.3}(Y_i \mid T_i = 1) - q_{.3}(Y_i \mid T_i = 0) \quad \text{; \underline{character}}$$

interpretation: the difference of 30% of population of receiving new drug & old drug.

$$\ln w_i^{(1)} = \beta_{0,\tau} + ed_i^{(1)} \beta_{1,\tau} + exp_i^{(1)} \beta_{2,\tau} + \varepsilon_{\tau,i}^{(1)} + \alpha_i$$

$$\ln w_i^{(2)} = \beta_{0,\tau} + ed_i^{(2)} \beta_{1,\tau} + exp_i^{(2)} \beta_{2,\tau} + \varepsilon_{\tau,i}^{(2)} + \alpha_i$$

$$q_\tau(\ln w_i^{(1)} \mid ed_i^{(1)}, exp_i^{(1)}, \alpha_i) = \beta_{0\tau} + ed_i^{(1)} \beta_{1\tau} + exp_i^{(1)} \beta_{2,\tau}^{(1)} + \alpha_i$$

$$\Delta \ln w_i = \Delta\, ed_i \beta_{1,\tau} + \Delta\, exp_i \beta_{2,\tau} + \Delta\, \varepsilon_{\tau,i}$$

$$q_\tau(Y_{1,i}) - q_\tau(Y_{0,i})$$

QTE

3. $y_{i,t} = \beta_0 + X_{i,t}\beta_1 + \alpha_i + \varepsilon_{i,t}$ $\alpha_i \to + \gamma_i$

① i.i.d

② $E(U_{i,t} \mid x_{i,t}, x_{t+\alpha}) = 0$

(a) $\Delta y_i = y_{i,3} - y_{i,1}$

$\Delta x_i = x_{i,3} - x_{i,1}$

$$\hat{\beta}_{OLS} = \frac{\frac{1}{n}\sum x_i \Delta Y_i}{\frac{1}{n} \Delta X_i^2} \xrightarrow[p]{WLN} \frac{E(\Delta X_i \Delta Y_i)}{E(\Delta X_i^2)} = \beta_1,$$

$P_r(\Delta X_i = 0) < 1$ $\xrightarrow{avoid}$ $\Delta X_i = 0$ always

$\Delta X_i \neq 0$

$$\Delta y_i = y_{i,3} - y_{i,1} = \underbrace{(X_{i,3} - X_{i,1})}_{\Delta X_i}\beta_1 + \underbrace{\mu_{i,3} - \mu_{i,1}}_{E\,\Delta\mu_i}$$

$$E(\mu_{i,3}\Delta X_i) = E(E(\mu_{i,3}\Delta X_i \mid \Delta X_i)$$

$$+ \frac{E(\Delta X_i 2\gamma_i)}{E(\Delta X_i^2)} \neq 0$$

$$\hat{\beta}_{OLS} = \begin{pmatrix} \sum 1 & \sum \tilde{X}_i \\ \sum \tilde{X}_i & \sum \tilde{X}_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum \tilde{y}_i \\ \sum \tilde{X}_i \tilde{y}_i \end{pmatrix}$$

$\xrightarrow{p} \theta$

PS5. 2.

$X_1 = t_1$, $X_2 = t_2$, $X_3 = t_3$ n sample

$n_1 \to \hat{q}_1$, $n_2 \to \hat{q}_2$, $n_3 \to \hat{q}_3$

$$\beta_1 = \frac{\sum (X_i - \bar{X})(y_i - \bar{y})}{\sum (X_j - \bar{X})^2} \xrightarrow{p}$$

$t_3 - t_2 = t_2 - t_1 = h$

TP: $\hat{q}_i \xrightarrow[q]{} \beta_0 + \beta_1 z_i$