Econ 706

Overview: Treatment Effects $\begin{cases} Y_1 \leftarrow \\ Y_0 \leftarrow \end{cases}$ $E(Y_1 | T=1) = E(Y_1)$

drug treat

Treatment Effects

Notation

i subscripts on random variables to be clear that this is individual specific.

$E(Y_i | X_i = x)$ function of x

$\downarrow$ number

E

key theme - causal estimation

estimate the effect of some intervention or policy on an individual

drug

$$Y_i = \boxed{Y_{0i} \quad Y_{1i}}_{T_i}$$

counterfactual

treatment effect for individual i

$$\Delta_i = Y_{1i} - Y_{0i}$$

☆ A feature of distribution

★ Average Treatment Effect (ATE) $E(\Delta_i)$

☆ Treatment on the Treated (TT) $E(\Delta_i | T_i = 1)$

Treatment on the Untreated (TUT) $E(\Delta_i | T_i = 0)$

more interesting

Identification

different things to different people

~~model~~ is uniquely determined by the population version of the data

$\theta \Rightarrow \begin{cases} ATE & \Rightarrow \text{homogenious} \\ TT \\ TUT \end{cases}$

$\Delta_i \Rightarrow$ hetergenious.

(Y_{0i}, Y_{1i}, T_i)

(Y_i, T_i) $\Rightarrow$ We can observe.
sample version (X population)

in the population
θ is identified

We can observe
the distribution of Y_{0i} conditional on $T_i = 0$
. . . . . . Y_{1i} . . . . $T_i = 1$

$$TT = E(\Delta_i \mid T_i = 1) = E(Y_{1i} - Y_{0i} \mid T_i = 1)$$
$$= \underbrace{E(Y_{1i} \mid T_i = 1)}_{\text{identified}} - \underbrace{E(Y_{0i} \mid T_i = 1)}_{\text{counterfactual}}$$

How to deal with

Sep 10th.
Randomized Controls Trials

$$\Pr(Y_i = y_j \mid X_i = x_i) = \frac{\Pr(Y_i = y_i, X_i = x_i)}{\Pr(X_i = x_i)}$$
$$= \frac{\Pr(Y_i = y_j, X_i = x_i)}{\sum_{j=1}^{k_y} \Pr(Y_i = y_j, X_i = x_i)}$$

X_i and Y_i are independent then
$$E(Y_i \mid X_i = x_i) = E(Y_i)$$

X_i is independent of (Y_i, Z_i)
$$E(Y_i \mid X_i = x_i, Z_i = z_m) = E(Y_i \mid Z_i = z_m)$$

LIE $\quad E(E(Y_i \mid X_i)) = E(Y_i)$

$$E(E(Y_i \mid X_i, Z_i) \mid Z_i = z_m)$$
$$= E(Y_i \mid Z_i = z_m)$$

Bayes Theorem
$$\Pr(X_i = x_i \mid Y_i = y_i)$$

⇒ randomization ⇒ framework ⇒ identical

∴ Selective attrition

Typically in an

A $\quad$ ~~B~~

$E(\Delta \mid A)$ ✓

if from original population ✗

Randomization

$E(\Delta \mid A)$

$E(\Delta \mid B)$

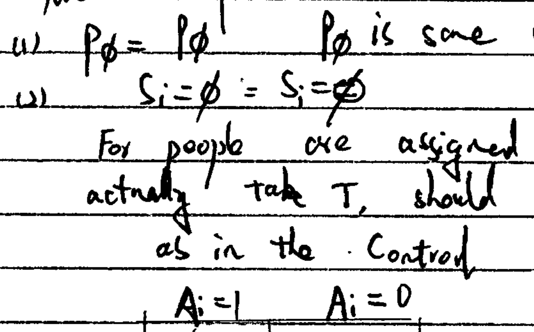
$E(Y_{1i} \mid \spadesuit, \diamond) - E(Y_{0i} \mid \spadesuit)$

Two assumptions:

(1) $P_\phi = P_\phi'$ P_ϕ is same in T and C

(2) $S_i = \phi = S_i = \phi$

For people are assigned in T but not actually take T, should be identically treated as in the Control group (sometime false,



e.g. Tell people in T if you exercise regularly with drug you'll be OK. But the people don't take drug but exercise, then this assumption (2) X

we can observe A, B, C.

but in C, you know B=2 (not observable)

so. $(\tau - \phi)$

Sample Example

1. Problem. Because the government announce the plan ahead, then people will extend their time in 1A ~~and~~ in order to get the subsidy.

Balancing Test (Table II)

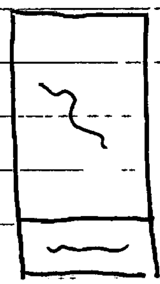
Example 2

3 group

- control
- section 8 group — anywhere — CP / NCP
- experimental group — poverty rate $<10\%$ — CP / NCP

CP: compliers (uses the voucher to move)

Sep 22nd.



Sep 24th.

List

social experiments.

lab experiments - prospect theory - endowments matter

Selection only on Observation

Assumption 1:

$$E(Y_{1i} \mid X_i = x, T_i = t) = E(Y_{1i} \mid X_i = x)$$

$$E(Y_{0i} \mid X_i = x, T_i = t) = E(Y_{0i} \mid X_i = x)$$

unconfoundedness

Assumption 2

$$Pr(T_i = 0 \mid X_i = x) > 0$$

Assumption 3

$Pr(T_i = 1 \mid X_i = x_i) > 0$ → this is harder

TUT → Assumption 1, 3

IT → Assumption 2

Sep 29th

The distribution is same

We need additional assumptions

A2 $Pr(T_i = 0 \mid X_i = x) > 0$

A3 $Pr(T_i = 1 \mid X_i = x) > 0.$

A3 is stronger than A2.

Theorem 1

Under A1 A2 TT is identified.

Under A3

$E(Y_{1i} \mid T_i = 0) = \sum_k E(Y_{1i} \mid X_i = x_j, T_i = 1) \, Pr(X_i = x_i \mid T_i = 0)$

		C	T
$\frac{1}{2}$	A	90%	10%
$\frac{1}{2}$	B	1%	99%

Under A3

Estimation.

Allowing for heterogeneous treatment effects in

$$Y_{0i} = X_i'\beta_0 + \mu_{0i}$$
$$Y_{1i} = X_i'\beta_1 + \mu_{1i}$$

$$\widehat{ATE} = \frac{1}{N}\sum_{i=1}^{N} X_i'(\hat{\beta}_1 - \hat{\beta}_0)$$

$$X_i'\hat{\beta}_0 \approx E(Y_{0i}|X_i)$$

$$\widehat{ATE} = \frac{1}{N}\sum_{i=1}^{N} T_i[Y_i - X_i'\hat{\beta}_0] + (1-T_i)[X_i'\hat{\beta}_1 - Y_{0i}]$$

Matching

Step 1

$$\widehat{TT} = \frac{1}{N_1}\sum_{(i:T_i=1)} Y_i - Y_{J_0(i)}$$

Propensity Score Matching

$$P_{(x)} \equiv Pr(T_i=1|X_i=x)$$

lower dimensional

$$p(x) \equiv Pr(T_i=1|X_i=x)$$

$$F(x|P(X_i)=p, T_i=1)$$
$$= F(x|P(X_i)=p)$$

$$F(x|P(X_i)=p, T_i=0) = F(x|P(x)=p, T_i=1)$$

$$E(Y_{0i}|\text{Sample})$$

$$E(Y_{1i}|\text{Sample})$$

propensity draw some distribution from sample.

~~asympto~~ asymptotic

Kernel

$$E(Y_{0i} \mid T_i = 1,\ P(x_i) = p)$$
$$= \int E(Y_{0i} \mid X_i = x)\, dF(x \mid T_i = 1,\ P(X_i) = p)$$
valid AI $\rightarrow$ $$= \int E(Y_{0i} \mid X_i = x)\, dF(x \mid T_i = 0,\ P(X_i) = p)$$
$$= E(Y_{0i} \mid T_i = 0,\ P(X_i) = p)$$

distribution of $P(X_i) = p$ is different,
because the treatment is ~~more~~ easier to get treatment.

$$P(X_{I_0(i)}) = P(X_i)$$
$$T_{I_0(i)} = 0$$

$$E(\widehat{TT}) = \int E(Y_i - Y_{I_0(i)} \mid T_i = 0, P(X_i) = p)\, f(p \mid T_i = 1)\, dp$$
...
$$= E(Y_{1i} - Y_{0i} \mid T_i = 1)$$

$\Delta_{41} = Y_{i4} - Y_{i1}$ Men Women

Using
$\downarrow$
Endogeneity

Does Piped Water Reduce Diarrhea for children in Rural India

Generalized Method of Moments
moment conditions
$$E(g(x_i, \theta_0)) = 0$$

N

$$f(\hat\theta) = f(\theta_0) + \frac{\partial f(\hat\theta)}{\partial \theta'}(\hat\theta - \theta_0)$$

$$= f(\theta_0) + \frac{\partial f(\theta_0)}{\partial \theta'}(\hat\theta - \theta_0)$$

$$\sqrt{N}\,(f(\hat\theta) - f(\theta_0)] \approx \underbrace{\frac{\partial f(\theta_0)}{\partial \theta'}}_{\text{scalar}} \sqrt{N}(\hat\theta - \theta_0)$$

$$\sim N\left(0, \frac{\partial f(\theta_0)}{\partial \theta'}\,\phi\,\frac{\partial f(\theta_0)}{\partial \theta}\right)$$

$$f(\theta) = \frac{1}{N}\sum_{i=1}^{N} g(x_i, \theta)$$

$$\frac{1}{N}\sum_{i=1}^{N} g(x_i, \hat\theta) = \frac{1}{N}\sum_{i=1}^{N} g(x_i, \theta_0) + \frac{1}{N}\sum_{i=1}^{N}\frac{\partial g(x_i, \hat\theta)}{\partial \theta'}(\hat\theta - \theta_0)$$

$$\sqrt{N}\Big(\underbrace{\frac{1}{N}\sum_{i=1}^{N} g(x_i, \hat\theta)}_{0} - \underbrace{\frac{1}{N}\sum_{i=1}^{N} g(x_i, \theta_0)}_{0}\Big) = \xrightarrow{\text{cd.}} \text{Normal distribution}$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial g(x_i, \hat\theta)}{\partial \theta'}\left[\sqrt{N}\,(\hat\theta - \theta_0)\right]$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial g(x_i, \hat\theta)}{\partial \theta'} \approx \frac{1}{N}\sum_{i=1}^{N}\frac{\partial g(x_i, \theta_0)}{\partial \theta'}$$

$$\approx E\left(\frac{\partial g(x_i, \theta_0)}{\partial \theta'}\right)$$

$$\equiv G$$

$$\sqrt{N}(\hat\theta - \theta_0) = -\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial g(x_i, \hat\theta)}{\partial \theta'}\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g(x_i, \theta_0)$$

$$\approx -G^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g(x_i, \theta_0)$$

$$\approx N(0, G^{-1}\psi G'^{-1})$$

iid.

in OLS: $g(x_i, \beta) = X_i[Y_i - X_i'\beta]$

$$G = E\left(\frac{\partial g(x_i, \theta_0)}{\partial \theta'}\right) = E(X_i X_i')$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} X_i X_i'$$

$$\psi = E\left[g(x_i, \theta_0)\, g(x_i, \theta_0)'\right]$$

MLE

$$\sum_{i=1}^{N} \frac{\partial l(x_i, \hat{\theta})}{\partial \theta} = 0.$$

$$G = E\left(\frac{\partial g(x_i, \theta_0)}{\partial \theta'}\right)$$

$$= E\left(\frac{\partial^2 l(x_i, \theta_0)}{\partial \theta \partial \theta'}\right)$$

$$\psi = E\left[g(x_i, \theta_0)\, g(x_i, \theta_0)'\right]$$

$$= E\left[\frac{\partial l(x_i, \theta_0)}{\partial \theta} \frac{\partial l(x_i, \theta_0)}{\partial \theta'}\right]$$

Overidentified Case

$$E(g(x_i, \theta_0)) = 0.$$

$$\frac{1}{N} \sum_{i=1}^{N} g(x_i, \hat{\theta}) = 0.$$

$$\left[\frac{1}{N} \sum_{i=1}^{N} g(x_i, \theta)\right]' W \left[\frac{1}{N} \sum_{i=1}^{N} g(x_i, \theta)\right]$$

$$E(Y_i \mid X_i = x) = g(x_i \theta)$$

$$E(Y_i - g(x_i, \theta_0) \mid X_i) = 0$$

10. 13th

$$E(Y_i \mid X_i = x_i) = x'\beta$$

$$X_i (Y - X_i \beta)$$

$$Y_i = X_i \beta + u_i$$

$$E(u_i \mid X_i) = 0.$$

Binary Dependent Variables

$E(D_i \mid X_i = x) = 1 \Pr(D_i = 1 \mid X_i = x) + 0 \Pr(D_i = 0 \mid X_i = x)$

$= \Pr(D_i = 1 \mid X_i = x)$

$E(D_i \mid X_i = x) = G(x'\beta) = \Pr(D_i = 1 \mid X_i = x)$

e.g. $G(x'\beta) = \Phi(x'\beta)$

$G(x'\beta) = \frac{e^{x'\beta}}{1+e^{x'\beta}}$ → close form.

$f(x_i)(D_i - G(x_i'\beta))$

↓

$G(x_i'\beta)$ • $D_i = 1$

$1 - G(x_i'\beta)$ if $D_i = 0$

$\frac{1}{N}\sum_{i=1}^{N}[D_i \log (G(x_i'\beta)$

$$f(x_i) = \frac{g(x_i'\beta) X_i}{G(x_i'\beta)(1 - G(x_i'\beta))}$$

Panel data

N units (people, countries, firms)

$Y_{it} = X_{it}'\beta + \mu_{it}$

$E(\mu_{it} X_{it}) = 0$

standard error

Assumption $\operatorname{cov}(\mu_{it}\ \mu_{jt}) = 0$ ✓

Assumption $\operatorname{cov}(\mu_{it}, \mu_{i\tau}) = 0$ X

$$G = E\left(\frac{\partial g(X_{it}, \theta_0)}{\partial \theta'}\right)$$

$\Sigma_{i=1}^{N}\ \Sigma$

$\mathrm{cov}(\varepsilon_{i1}, \varepsilon_{i2}) = 0$

$$\hat{\beta}_0 = \frac{1}{2N} \sum_{i=1}^{N} \sum_{t=1}^{2} Y_{it}$$

$$\sqrt{2N}\,(\hat{\beta}_0 - \beta_0) \sim N(0, V)$$

$$V = \mathrm{Var}(Y_{it}) \approx \frac{1}{2N} \sum_{i=1}^{N} \sum_{t=1}^{2} (Y_{it} - \hat{\beta}_0)^2$$

GMM $\quad G($

GMM formula. $\quad \frac{\partial G(Y_{it}, \beta_0)}{\partial \beta_0} = -1$

$$V = E(G(Y_{it}, \beta_0)^2)$$

$$\sqrt{2N}\,(\hat{\beta}_0 - \beta_0) \sim N(0, \quad V$$

$$V \approx \frac{1}{2N} \sum_{i=1}^{N} \sum_{t=1}^{2} (Y_{it} - \hat{\beta}_0)^2$$

panel data.

$$g(X_{it}, \theta_0)$$

$$\gamma(X_{it}, \theta_0) = \sum_{t=1}^{T_i} g(X_{it}, \theta_0)$$

$$E(\gamma(X_{it}, \theta_0)) = 0.$$

$$0 = \frac{1}{N} \sum_{i=1}^{N} \gamma(X_i, \hat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T_i} g(X_{it}, \hat{\theta})$$

Weight equally

$$G = E\left(\frac{\partial \gamma(X_i, \theta_0)}{\partial \theta}\right)$$

Variance $V = E\left(\gamma(X_i, \theta_0)\,\gamma(X_i, \theta_0)'\right)$

$\gamma(Y_i, \beta_0)$ assume $cov(\mu_{it_1}, \mu_{it_2}) = 0$

不同人之间是独立的

$$\gamma(Y_i, \beta_0) = \sum_{t=1}^{2} (Y_{it} - \beta_0)$$

$$V_0 = E(\gamma(Y_i, \beta_0)^2)$$
$$= E\left(\left(\sum_{t=1}^{2} (Y_{it} - \beta_0)\right)^2\right)$$
$$= E\left(\left(\sum_{t=1}^{2}(\theta_i + \varepsilon_{it})\right)^2\right)$$
$$= \sum_{t_1=1}^{2} \sum_{t_2=1}^{2} E[(\theta_i + \varepsilon_{i1})]^2$$

$$\sqrt{N}(\hat{\beta}_0 - \beta_0) \sim N(0, G^{-1} V G^{-1})$$
$$Var(\hat{\beta}_0) \approx \frac{V}{NG^2} = 4\sigma_\theta^2 + 2\sigma_\varepsilon^2$$

$$V = E(\gamma(X_i, \theta_0)\,\gamma(X_i, \theta_0)')$$

reg y, x, cluster(i)

$$\sigma_{it\tau} = cov(\mu_{it}, \mu_{i\tau})$$

asyption should be there and.

$\sum\sum\sum$

gen

reg

reg

reg

xtreg

re(id)

Fixed Effects

$$Y_{it} = X_{it}\beta + \theta_i + \varepsilon_{it}$$

$$E(\theta_i X_{it}) = 0$$

don't need X

$$\overline{Z_i} \equiv \frac{1}{T_i}\sum_{i=1}^{N} Z_{it}$$

$$E[(\varepsilon_{it} - \overline{\varepsilon}_i)(X_{it} - \overline{X}_i)] = 0$$

D_{it} $N \times 1$ vector

$$D_{it}^{(g)} = \begin{cases} 1 & i=j \\ 0 & \text{otherwise} \end{cases}$$

logit / probit
probability of OLS

Same. $Y_{it} = X_{it}\hat{\beta} + D_{it}'\hat{\delta} + \hat{\mu}_{it}$

consistent — estimator of θ_i
not consistent

but numeric equovalient
special for OLS

Y_{it}

Oct. 21st
Model vs. Estimator

$$Y_{it} = X_{it}\beta + D_{it}\theta + \mu_{it}$$

dependents upone the sample.

estimate and model are seperate issues

Fixed effects

First Differencing

$$Y_{it} - Y_{it-1} = (X_{it} - X_{it-1})'\beta + \varepsilon_{it} - \varepsilon_{it-1}$$

$\checkmark\ n \times (T-1)$

standard fixed

$$Y_{i2} - \bar{Y}_i = Y_{i2} - \frac{Y_{i1}+Y_{i2}}{2}$$

$$= \frac{Y_{i2}-Y_{i1}}{2} \quad \checkmark \quad n \times t-1)$$

which one is better?

$$NT - N - K = N(T-1) - K$$

Only two period data, it's same.

$Y_{i2} - \bar{Y}_i$

More than two period, it's different.

Assume.

$$T_{it} = \begin{cases} 0 & t \leq \tau \\ 1 & t > \tau \end{cases}$$

begin $\tau+1$

$$\hat{\alpha}_{FE} = \text{scov}((T_{it} - \bar{T}_i), ($$

$\bar{\varepsilon}_A - \bar{\varepsilon}_B$

$\bar{\varepsilon}_\tau - \bar{\varepsilon}_{\tau-1}$

$\varepsilon_{it} = \varepsilon_{it-1} + \delta_{it}$ random walk.

$\varepsilon_{i2} - \varepsilon_{i1} = \beta_{i2}$

$$\frac{\varepsilon_{i2} + \varepsilon_{i3}}{2} - \varepsilon_{i1} = \frac{{}^{2}\varepsilon_{i2} + \varepsilon_{i3}}{2} - \varepsilon_{i1}$$

$$= \frac{2\xi_{i2} + \beta_{i3}}{2}$$

Fixed Effects Versus Regression

$$\sum_i \sum_t (X_{it} - \bar{X}_i)^2 + \sum_i T_i (\bar{X}_i - \bar{X})^2$$

within variance — between variance

FE: use within variance

OLS: use both.

for example, education (X_{it}), it is fixed.

so. OLS is much better.

$$\beta_{OLS} = \beta + \frac{\mathrm{cov}(X_{it}, \theta_i + \varepsilon_{it})}{\mathrm{var}(X_{it})}$$

$$\beta_{FE} = \beta + \frac{\mathrm{cov}(X_{it} - \bar{X}_i, \varepsilon_{it} - \bar{\varepsilon}_i)}{\mathrm{var}(X_{it} - \bar{X}_i)}$$

generally, $\mathrm{cov}(X_{it}, \theta_i + \varepsilon_{it}) < \mathrm{cov}(X_{it} - \bar{X}_i, \varepsilon_{it} - \bar{\varepsilon}_i)$

denominators: FE < OLS.

age. ~~exp~~ ~~cohort~~ current year

$\Downarrow$

multicollinearity.

non balanced data is OK

Solution: use twins

☆ Difference-in-Differences

Assume δ_t is same for two groups

$g(i)$ indicate and individual's group (either $\Diamond$ or $\clubsuit$)

$$T_{it} = \begin{cases} 0 & g(i) = \clubsuit, \; t \in \{0,1\} \\ 0 & g(i) = \Diamond, \; t=0 \\ 1 & g(i) = \Diamond, \; t=1 \end{cases}$$

Identification:

$\hat{\alpha} = (\bar{Y}_{\Diamond 1} -$

$$Y_{it} = \beta_0 + \alpha T_{it} + \delta_t + \theta_i + \varepsilon_{it}$$

$Cov(X, Y)$ _

$$Y_i = \beta_0 + \alpha T_{g(i)\,t(i)} + \delta t(i) + \gamma \Diamond(i) + \varepsilon_i$$

Category	$T_{g(i)\,t(i)}$	$t(i)$	$\Diamond(i)$
$\Diamond$ 0	0	0	1
$\Diamond$ 1	1	1	1
$\clubsuit$ 0	0	0	0
$\clubsuit$ 1	0	1	0

$$0 = \sum_{i=1}^{N} \hat{\varepsilon}_i = \sum_{\Diamond 0} \hat{\varepsilon}_i + \sum_{\Diamond 1} \hat{\varepsilon}_i + \sum_{\clubsuit 0} \hat{\varepsilon}_i + \sum_{\clubsuit 1} \hat{\varepsilon}_i$$

T_i: $\quad 0 = \sum_{i=1}^{N} \sum_{t=0}^{1} T_i \hat{\varepsilon}_i = \sum_{\Diamond 1} \hat{\varepsilon}_i$

t_i: $\quad 0 = \sum_{i=1}^{N} t_i \hat{\varepsilon}_i = \sum_{\Diamond 1} \hat{\varepsilon}_i + \sum_{\clubsuit 1} \hat{\varepsilon}_i$

$$\Rightarrow 0 = \sum_{\Diamond 0} \hat{\varepsilon}_i = \sum_{\Diamond 1} \hat{\varepsilon}_i = \sum_{\clubsuit}$$

$$Y_{it} = \beta_0 + \alpha T_{it} + \underbrace{\delta_t + \theta_i}_{\text{paramenter fixed effect.}} + \varepsilon_{it}$$

$$\text{Prob}(Y_i = 1) = \Phi(\beta_0 + \alpha T_{g(i)t(i)} + X_i'\beta + \delta_{t(i)} + \theta_{g(i)})$$

large sample it will be ok

( ) probit estimator

( ) st error

( ) $\frac{dP_r}{\partial x}$

DD see p lecture note

Instrumental Variables

$$E(T_i \mu_i) = 0 \qquad E(X_i \mu_i) = 0$$

$$Z_i^* = \begin{bmatrix} Z_i \\ X_i \end{bmatrix} \quad X_i^* = \begin{bmatrix} T_i \\ X_i \end{bmatrix} \quad \beta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$E(Z_i^* X_i^{*\prime}) \rightarrow$ randomly draw Z_i^* from data

Partitioned Regression

$$\underset{n\times n}{Y} = \underset{n\times n}{X_1}\beta_1 + \underset{n\times k_2}{X_2}\underset{k_2\times n}{\beta_2} + \mu$$

$$M_2 = I - \underset{x_2 \times k_2}{X_2}(X_2'X_2)^{-1}\underset{k_2\times n}{X_2'}$$

Fact1 M_2 is idempotent

Fact2 $M_2 Y$ is the Residuals from Regression

$Y = (Y_1, Y_2, Y_3)$

$m\ Y = ($

$\Rightarrow Y_1$ on X_1

Y_2 on X_2

Y_3 on X_3

$$M_2 X_2 = X_2 - X_2 (X_2' X_2)^{-1} X_2' X_2 = 0$$

GMM equations

$$0 = X_1' (Y - X_1 \hat{\beta}_1 - X_2 \hat{\beta}_2)$$

$$0 = X_2' (Y - X_1 \hat{\beta}_1 - X_2 \hat{\beta}_2)$$

$$\beta_2 = (X_2' X_2)^{-1} X_2' (Y - X_1 \hat{\beta}_1)$$

$$0 = X_1' (Y - X_1 \hat{\beta}_1 - X_2 \hat{\beta}_2)$$

$$= X_1' (Y - X_1 \hat{\beta}_1 - X_2 (X_2' X_2)^{-1} X_2' (Y - X_1 \hat{\beta}_1))$$

$$= X_1' M_2 Y - X_1' M_2 X_1 \hat{\beta}_1$$

$$\hat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 Y$$

$$= (\tilde{X}_1 \tilde{X}_1)^{-1} \tilde{X}' \tilde{Y}$$ numerically equivalent

IV

$$0 = Z' (Y - T \hat{\alpha}_{IV} - X \hat{\beta}_{IV})$$

$$0 = X' (Y - T \hat{\alpha}_{IV} - X \hat{\beta}_{IV})$$

$$\hat{\beta}_{IV} = (X'X)^{-1} X' (Y - T \hat{\alpha}_{IV})$$

$$\hat{\alpha}_{IV} = (Z' M_X T)^{-1} Z' M_X Y$$

$$= \frac{\tilde{Z}' \tilde{Y}}{\tilde{Z}' \tilde{T}} \approx \frac{\mathrm{cov}(\tilde{Z}_i, \tilde{Y}_i)}{\mathrm{cov}(\tilde{Z}_i, \tilde{Y}_i)}$$

mean = 0.

$$\tilde{Y} = M_X Y = \alpha M_X T + M_X X \beta + M_X u$$

$$= \alpha \tilde{T} + \tilde{u}$$

$$\hat{\alpha}_{IV} \approx \frac{\mathrm{cov}(\tilde{Z}_i, \tilde{Y}_i)}{\mathrm{cov}(\tilde{Z}_i, \tilde{T}_i)} \approx \frac{\mathrm{cov}(\tilde{Z}_i, \alpha \tilde{T}_i + \tilde{u}_i)}{\mathrm{cov}(\tilde{Z}_i, \tilde{T}_i)}$$

$$= \alpha + \frac{\mathrm{cov}(\tilde{Z}_i, \tilde{u})}{\mathrm{cov}(\tilde{Z}_i, \tilde{T}_i)}$$

$\mathrm{cov}(\tilde{Z}_i, \tilde{u}_i) = 0$
$\mathrm{cov}(\tilde{Z}_i, \tilde{T}_i) \neq 0$

$$\hat{\alpha}_{OLS} \approx \alpha + \frac{\mathrm{cov}(\tilde{T}_i, \tilde{u}_i)}{\mathrm{cov}(\tilde{T}_i, \tilde{T}_i)}$$

$\hat{\alpha}_{OLS} = 0.1$ biased, $\Rightarrow$ between 0.1 0.5
$\hat{\alpha}_{IV} = 0.5$

Simultaneous equations

$Y_i = \alpha T_i + X_i \beta + u_i$
$T_i = \rho Y_i + X_i \gamma + Z_i \delta + v_i$ structural equations

T_i & u_i may be correlated.

$$\beta_1^* = \frac{\alpha \gamma + \beta}{1 - \alpha \rho}$$

$$\delta_1^* = \frac{\alpha \delta}{1 - \rho \alpha} \qquad \delta_2^* = \frac{\delta}{1 - \rho \alpha}$$

we need to ~~assume the~~ existence of Z

$$\frac{\delta_1^*}{\delta_2^*} = \alpha$$

$$\begin{tikzcd} Z_i \arrow[r, "\delta_2^*"] & T_i \arrow[r, "\alpha"] & Y_i \end{tikzcd}$$

$$\begin{tikzcd} & \tau_i \arrow[dr, "a"] & \\ Z_i \arrow[ur, "\delta"] \arrow[dr, "\delta_2^*"'] & & Y_i \\ & T_i \arrow[ur, "\alpha"'] & \end{tikzcd}$$

return of e.g. schooling

T_i quantity
τ_i quality

one IV

$$\hat{\delta}_1 = \frac{\tilde{Z}_i' Y_i}{\tilde{Z}_i' Z_i} \qquad \hat{\delta}_2 = \frac{\tilde{Z}_i' \tilde{T}_i}{\tilde{Z}_i' Z_i}$$

$$\frac{\hat{\delta}_1}{\hat{\delta}_2} = \frac{\tilde{Z}_i' Y_i}{\tilde{Z}_i' \tilde{T}_i} = \hat{\alpha}_{IV}$$

more than 1 Z

Method 2.

$$T_i^{\dagger} = X_i' \beta_2^* + Z_i' \delta_2^*$$

X_i' is good variation ⇒ uncorrelated μ_i the Z_i part

but cannot seperate T_i and X_i

$$\begin{aligned} Y_i &= \alpha T_i + X_i'\beta + \mu_i \\ &= \alpha [T_i^{\dagger} + V_i^*] + X_i'\beta + \mu_i \\ &= \alpha T_i^{\dagger} + X_i'\beta_2 + (\alpha V_i^* + \mu_i) \end{aligned}$$

⇓ consistent

$$T_i^{\dagger} = X_i'\beta_2^* + Z_i'\delta_2^*$$

so X perfect multicolinearity otherwise

Two - Stage Least Square

$$T_i = \rho Y_i + X_i'\gamma + Z_i\delta + V_i \quad \Rightarrow \text{GMM}$$

$$Y_i = \alpha T_i + X_i'\beta + \mu_i \quad \Rightarrow X$$

We need to get consistent estimates

$$Z^* = (Z, X)$$
$$X^* = (T, X)$$

$$\hat{T} = Z^* (Z^{*\prime} Z^*)^{-1} Z^* X^*$$

3 Interpretations

① GMM estimator or $(Z'T)^{-1} Z'Y$

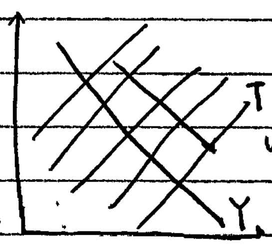
② Ratio of reduced form estimates - rescaling the reduced form

③ 2SLS - direct effect of fitted model

e.g. ① Simultaneity bias : $\rho \neq 0$ ☆

② u_i correlated with T_i

③ Measurement Error



use Z_i to shift T_i to trace Y

A classic (perhaps the classic) example is the returns to schooling!

If IV estimates are higher than OLS

$$\hat{\alpha} \approx \frac{Cov(Z_i, Y_i)}{cov(Z_i, T_i)} = \frac{cov(Z_i, \varepsilon_i + \Delta_i T_i)}{Cov(Z_i, T_i)}$$

$$= \frac{Cov(Z_i, \Delta_i T_i)}{cov(Z_i, T_i)} = \alpha$$

$$\Delta_i = \alpha$$

Imbens and Angrist (1994)

$$E(\Delta_i T_i Z_i) - E(\Delta_i T_i)E(Z_i)$$

$$= p_z(1-p_z)\mu_\heartsuit E(\Delta_i = 1 \mid S_i = \heartsuit)$$

$$\hat{\alpha} \approx \frac{p_z(1-p_z)\mu_\heartsuit E(\Delta_i \mid S_i = \heartsuit)}{p_z(1-p_z)\mu_\heartsuit}$$

$$= E(\Delta_i \mid S_i = \heartsuit)$$

11.19

$$\hat{\alpha} \approx \frac{P_2(1-P_2)\mu_\heartsuit E(\Delta_i \mid S_i = \heartsuit)}{P_2(1-P_2)\mu_\heartsuit}$$

$$= E(\Delta_i \mid S_i = \heartsuit)$$

local average treatment effect (LATE)

This is in fact IV

O_i as an instrument for T_i

in fact the TT estimator is the LATE for the IV estimator

D_i $\quad \hat{\alpha} = \frac{\hat{\delta}_1^*}{\hat{\delta}_2^*} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{T}_1 - \bar{T}_0}$

$\bar{T}_0 = 0 \Rightarrow$ TT estimator

Complier Average Treatment Effect.

~~the~~ people will do as you think.

$\beta_1 S_i + \beta_2 S.T_i$

test score.

lower perform children get more advancement.

Paper 2 Acemoglu

R → protection against expropriation.

1 - 10

Measurement error $Y_i = \beta_0 + \alpha T_i + u_i$

OLS is higher than IV → measurement error

Overidentification

Z_i is multidimensional

$k_{z_i} > K_x$

$$W = E(u_i^2 Z_i^* Z_i^{*\prime})^{-1}$$

$$(E(X_i^* Z_i^{*\prime}) E(u_i^2 Z_i^* Z_i^{*\prime})^{-1} E(Z_i^* X_i^*))^{-1}$$

Tests

$$k+2 \begin{cases} \overbrace{0 = Z_1'(Y - T\hat{\alpha} - X\hat{\beta})}^{k+1} \\ 0 = Z_2'(Y - T\hat{\alpha} - X\hat{\beta}) \\ 0 = X'(Y - T\hat{\alpha} - X\hat{\beta}) \end{cases}$$

$$0 = Z^{*\prime}(Y - X^*\hat{\beta})$$

$$0 = Z_i'(\tilde{Y} - \tilde{T}\alpha)$$

$$\hat{\alpha}_1 \approx \hat{\alpha}_2$$

+ : IVs works on different populations

Regression Discontinuity

$$T_i = \begin{cases} 0 & X_i < x^* \\ 1 & x_i \geq x^* \end{cases}$$

$E(\varepsilon_i \mid x_i = x)$ is smooth

$$\lim_{x \uparrow x^*} E(Y_i \mid X_i = x) = E(\varepsilon_i \mid X_i = x^*)$$

$$\lim_{x \downarrow x^*} E(Y_i \mid X_i = x) = \alpha + E(\varepsilon_i \mid X_i = x^*)$$

$= \alpha$

around x^* T is randomly assinged

Fuzzy Regression Discontinuity

Wald Estimator

$E(\alpha \mid X = x^*)$

so it's very narrow set of people to estimate.
$X = x^*$

$x + \varepsilon$:

Kernel Regression.
$h \rightarrow$ put weight on x that is around x^*.

$h \uparrow$ put equal weight on all observation variance is $\downarrow$ bias $\uparrow$
$h \downarrow$ variance $\uparrow$ bias $\downarrow$

h is small

$$\frac{x_i - x}{h}$$

problem: overestimate
solution: run regression and forcast

Another approach.

$$g(x) = E(\varepsilon_i \mid X_i = x_0)$$

$$E(Y_i \mid X_i, T_i) = \alpha T_i + g(X_i)$$

$g(x)$ is a smooth function.

Problems.

① Sample is too sample ⇒ bandwidth is so large
the degree of the polynomial so small
isn't a regression discontinuity

② $g(x)$ is not smooth.
{ not only policy happened some other things use exact same cutoff.

③ running variable is endogenous

Note that you need X_i to be precisely manipulated, if there is still some randomness on the actual value of X_i rd looks fine.

7230

border fixed effects

Maimonides' Rule

Does Air Quality Matter? Evidence from the Housing Market.

$$d_i^{U_{ij}} = \underset{j \in (0..j-1)}{\arg\max}^{0 \dots j-1} \{U_{ij}\}$$

$$U_{ij} = a_j + X_i' \beta_j + Z_j' \delta + v_{ij}$$

from firm's perspective, we only care about the demand, but not people's happiness, so do not include the term with i

a_j how much people like _

β_j how people's characteristics influence the demand

δ how good's characteristics influence the demand
how much you like the good.

We could have $\underline{Q_{ij}' \delta_j}$

Identification in Binary Case

$j = 2$
cardinal

$j = 1$

$$U_{i1} > U_{i0}$$

$$\Leftrightarrow a_1 + X_i' \beta_1 + Z_1' \delta + v_{i1} > a_0 + X_i' \beta_0 + Z_0' \delta + v_{i0}$$

$$\Leftrightarrow \underbrace{(a_1 + Z_1' \delta - a_0 - Z_0' \delta) + X_i' (\beta_1 - \beta_0)}_{U_{i1}^*} + \underbrace{V_{i1} - V_{i0}}_{U_{i0}^*} > 0$$

normalization $a_0 = 0$ $\delta = 0$

$\beta_0 = 0$

$$(\vec{Z_1}' - \vec{Z_0}) \delta_1 + (Z^2 - Z_0^2) \delta_2$$

$a_0 = a_1 = 0$

$\beta_0 = 0.$

$v_{i0} = 0 \qquad V_{i0} = 0$

$v_{i1} = N(\mu, \sigma^2)$

$\mu = 0$

$a_1 + X_i' \beta_j + V_{ij} \geq 0$

$a^* + X_i' \beta_j^* + V_i \geq 0$

probit

$\frac{a_1 - a_0 + (z_i z_i) \delta}{\sigma} \qquad \frac{\beta_1 - \beta_0}{\sigma}$

$\Leftrightarrow$ a.

$\left(\frac{a_1 - a_0 + z_i \delta - z_i \delta}{\sigma}\right) + X_i' \left(\frac{\beta_1^* - \beta_0}{\sigma}\right) + \frac{V_{i1} - V_{i0}}{\sigma} \geq 0$

① ΔV logistic

$P_r(\Delta V < X) = \frac{e^x}{1+e^x}$

$\frac{\partial \Phi(x'\beta)}{\partial x} = \phi(x'\beta)\beta$

Probit $\cdot dx$

dProbit dx

②

run probit model

$$Pr(d_i = 1 \mid X_i = x)$$
$$= Pr(a_1 + X_i'\beta_1 + Z_i'\delta + v_{i1} > a_0 + X_i'\beta_0 + Z_i'\delta + v_{i0})$$

More than two choices

$$d_i = \begin{cases} 0 & U_{i0} > U_{i1}, \\ 1 & U \\ 2 & \end{cases}$$

still just need 1 normalization

Solution Multinomial Logit

$$U_{ij} = \mu_{ij} + v_{ij}$$

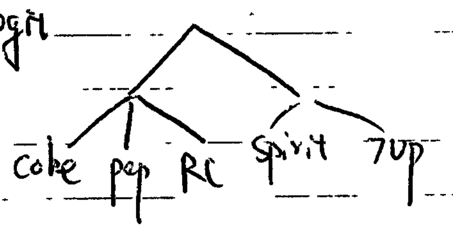
Substitution Patterns

$j=0$: car
$j=1$: Red bus
$j=2$: Blue bus

$$\frac{Pr(d_i=1)}{Pr(d_i=0)} = \frac{e^{\mu_{i1}}}{e^{\mu_{i2}}} \rightarrow \text{problem}$$

have nothing to do with μ_{i2}

Nest Logit



Mixed Logit

$$V_{ij} = a_j + X_i' \beta_{ij} + Z_j' \delta_i + v_{ij}$$

Sep 18th.

① Review of important concepts

→ underlying Counterfactual State

$$Y_i = \begin{cases} Y_{1i} & \text{when } T_i = 1 \\ Y_{0i} & T_i = 0 \end{cases}$$

Y_{1i}, Y_{0i} non-dependent

Treatment on the treated

$$E(Y_{1i} - Y_{0i}) = E(Y_{1i} - Y_{0i} \mid T_i = 1)\, Pr(T_i = 1)$$
$$+ E(Y_{1i} - Y_{0i} \mid T_i = 0)\, Pr(T_i = 0)$$
$$= [E(Y_{1i} \mid T_i = 1) - E(Y_{0i} \mid T_i = 1)]\, Pr(T_i = 1)$$
$$+ [E(Y_{0i} \mid T_i = 0) - E(Y_{0i} \mid T_i = 0)]\, Pr(T_i = 0)$$

↓ randomization

$$T_i \perp Y_{1i}, Y_{0i}$$

$$E(Y_{0i} \mid T_i = 1) = E(Y_{1i} \mid T_i = 0)$$

Issues.

① Selective Attrition.

② No perfect Compliance

↓

Difference between A_i and T_i

A_i: randomization; T_i: not random

2 types:

CP: $T_i = 1$ when $A_i = 1$

NCP: $T_i = 0$ regardless of A_i

What we could identify from data with random assignment.

$$E(Y_i \mid A_i = 1) \quad \& \quad E(Y_i \mid A_i = 0)$$

$A_i = 1$ ① A_i is not random ? ×

② $\dfrac{Pr(T_i = 1 \mid A_i = 0)}{Pr(A_i = 0)} = 0.$

P2 Q2

	$E(Y_0)$	$E(Y_1)$
A	1	1.5
B	0.5	1.5
C	1.5	1
D	1	1

$Pr(A) = Pr(B)$

$A_i \leftarrow$ assigned randomly

"A" & "B" $\rightarrow T_i = 1$

"C" & "D" $\rightarrow T_i = 0$

$E(Y_{1i} - Y_{0i} | T_i = 1)$

$= E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 1)$

$= E(Y_{1i} | T_i = 1, A_i = 1) - E(Y_{0i} | T_i = 1)$

$= E(Y_{1i}$

$\tau = (Y_i | A_i = 0, \text{type}) \quad A_i = 1$

	$E(Y_i \mid A_i = 0, \text{type})$	$E(Y_i \mid A_i = 1, \text{type})$
A	1	1.5
B	0.5	1.5
C	1.5	1.5
D	1	1

$E(Y_{1i} - Y_{0i} | T_i = 1) = E(Y_{1i} - Y_{0i} | \text{"A or B"})$

$= E(Y_{1i} - Y_{0i} | \text{"A"}) \dfrac{Pr(\text{"A"})}{Pr(A, B)} \dfrac{\frac{1}{4}}{\frac{1}{2}} + E(Y_{1i} - Y_{0i} | \text{"B"}) \dfrac{Pr(\text{"B"})}{Pr(A, B)}$

$= 0.5 \times \frac{1}{2} + 1 \times \frac{1}{2} = 0.75$

$E(Y_{1i} | A_i = 1) - E(Y_{0i} | A_i = 0)$

$= E(Y_i | A_i = 1, \text{"A" or "B" or "C" or "D"})$

$- E(Y_{0i} | A_i = 0, \text{"A" or "B" "C" or "D"})$

☆ $A_i \Rightarrow$ we observe
$T_i \Rightarrow$ we ✗ observe.

$$(1.5\times\tfrac{1}{4}+1.5\times\tfrac{1}{4}+1.5\times\tfrac{1}{4}+1\times\tfrac{1}{4})-$$
$$=(1\times\tfrac{1}{4}+0.5\times\tfrac{1}{4}+1.5\times\tfrac{1}{4}+1\times\tfrac{1}{4})$$
$$=0.375$$

$$P(CP) = \Pr(A \text{ or } B) = \Pr(T_i=1 \mid A_i=1) = \tfrac{1}{2}$$

Sep. 28th.

$E(Y_i \mid T_i=1) - E(Y_i \mid T_i=0)$
$= E(Y_{1i} \mid T_i=1) - E(Y_{0i} \mid T_i=0)$
$= E(Y_{1i}) - E(Y_{0i})$ $\quad T_i \perp Y_{1i}$ & Y_{0i}

reg Y $T \Rightarrow$ ATE ?
$Y_i = \alpha + \beta T_i + \varepsilon$
$E(Y_i \mid T_i=1) = \alpha+\beta$
$E(Y_i \mid T_i=0) = \alpha$
$\beta = E(Y_i \mid T_i=1) - E(Y_i \mid T_i=0)$

Q2(a) $CP \Rightarrow T_i = A_i$. $\quad$ reg Y $T \Rightarrow$ ATE
$A_i =$ ↓ observe in data: Y_i T_i

(b) observe in data: Y_i T_i type.
WT $\Rightarrow E(Y_{1i} - Y_{0i} \mid \text{type})$

(c) observe "A" "B" "C"
not observe "D"
observe data $T_i = A_i$
$E(Y_i \mid T_i=1, \text{observe}) - E(Y_i \mid T_i=0, \text{observe})$
↓
A B C

(d) selective attrition
"D" → never answer.
"A" & "B" → answer if $T_i=1$
no $T_i=0$.

Today ① matching Q2 → PS #2
② GMM

Goal: ① describe a rule how to create the match
② use matching to construct TT

	X_i	$T_i=1$ Y_i	i	1-1 match $I_0(i)$	i	$T_i=0$ Y_i
$\boxed{1}$	(0,1)	2	1	11	11	3
	(0,1)	1	2	14	14	1
$\boxed{2}$	(1,0)	1	9	3	3	3
		1	13	3	5	3
~~$\boxed{3}$~~	~~(2,1)~~					
$\boxed{3}$	(2,0)	2	7		6	0
		0	15		12	2
$\boxed{4}$	(1,1)	0	10		8	2

$$TT = \frac{1}{N_T}\sum_{i=1}^{N_T}(Y_i - Y_{I_0(i)})$$
$$= \frac{1}{7}(2+1+1+1+2-3-1-3-3-2-2)$$
$$= -1$$

$\boxed{2}$ avg of control

$$TT = \frac{1}{N_T}\sum_i (Y_i - Y_{I_0(i)})$$
$$= \frac{1}{7}(2+1+\dots$$

not necessarily identical

2
2
3
3
1
1

STATA: teffect psmatch

GMM:

[1] 5 useful technique

1. LLN
2. CLT
3. Mean Value Thm
4. Slutsky Thm
5. Delta Method

1. LLN

degenerated distribution sample mean $X_i \overset{iid}{\sim} F(\cdot)$

$\Leftarrow \frac{1}{N}\sum X_i \xrightarrow{P} E(X_i)$

2. CLT $\sqrt{N}\ \frac{1}{N}\sum(X_i\ E(X_i)) \xrightarrow{d} N(0, Var(X_i))$

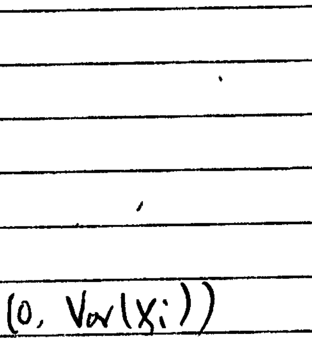
3. Mean Value Them

$f(\hat{\theta}) = f(\theta_0) + \frac{\partial f(\tilde{\theta})}{\partial \theta'}(\hat{\theta} - \theta_0)$

$\Leftarrow \exists \tilde{\theta}$ $\theta_0, \tilde{\theta}$

$f(\cdot)$ well behave

continous / smooth



4. Slutsky Theoem

$X_N \xrightarrow{P} X$

$Y_N \xrightarrow{P} Y$

$X_N \pm Y_N \xrightarrow{P} X \pm Y$

$X_N \cdot Y_N \xrightarrow{P} XY$

$\frac{X_N}{Y_N} \xrightarrow{P} \frac{X}{Y}$

5. Delta method.

$\sqrt{n}(\hat\theta - \theta_0) \xrightarrow{d} N(0, \phi)$

$\sqrt{n}(f(\hat\theta) - f(\theta_0)) \xrightarrow{d}$

$\hat\theta \xrightarrow{p} \theta_0 \quad \tilde\theta \xrightarrow{p} \theta_0$

$f(\hat\theta) \overset{MVT}{=} f(\theta_0) + \frac{\partial f(\tilde\theta)}{\partial \theta'}(\hat\theta - \theta_0) \leftarrow \tilde\theta$ in b/w θ_0, $\hat\theta$

$\sqrt{n}(f(\hat\theta) - f(\theta_0)) = \boxed{\frac{\partial f(\tilde\theta)}{\partial\theta'}} \sqrt{n}(\hat\theta - \theta_0)$

Slutsky Thm $\Rightarrow \underset{\text{constant}}{\frac{\partial f(\theta_0)}{\partial\theta'}} \underset{N\text{-}d}{\sqrt{n}(\hat\theta - \theta_0)}$

$\xrightarrow{d} N\left(0, \frac{\partial f(\theta_0)}{\partial\theta'} \phi \frac{\partial f(\theta_0)}{\partial\theta}\right)$

GMM

$E(g(X_i, \theta_0)) = 0 \overset{s/t.}{\Leftarrow} \exists!\ \theta_0$

OLS $\Leftarrow E[(Y_i - X_i\beta) X_i] = 0$

IV $\Leftarrow E[(Y_i - X_i\beta) Z_i] = 0$ $\quad \rightarrow \sqrt{n}(\hat\theta - \theta_0) \Leftarrow$ goal

MLE $\Leftarrow E \frac{\partial l(\theta_0)}{\partial\theta'} = 0$

Just Identified GMM

$E[g(x_i, \theta)] = 0$, L moment

K $\qquad$ K param

L

$L = K$

$\frac{1}{N}\sum_i g(x_i, \hat\theta) = 0$

$0 = \frac{1}{N}\sum g(x_i, \hat\theta)$

$= \frac{1}{N}\sum g(x_i, \theta_0) + \frac{1}{N}\sum \frac{\partial g(x_i, \tilde\theta)}{\partial\theta'}(\hat\theta - \theta_0)$

$\sqrt{n}(\hat\theta - \theta_0) = -\left[\frac{1}{N}\sum \frac{\partial g(x_i, \tilde\theta)}{\partial\theta}\right]^{-1} \frac{1}{N}\sum g(x_i, \theta_0)$

$\downarrow d$ $\qquad$ $\downarrow$ Slutsky $\qquad$ $\downarrow$ CLT

$N(0, G^{-1}\psi G^{-1})$

$\mathcal{M}_{it} = \theta_i + \varepsilon_{it}$ θ_i iid across
ε_{it} iid across i/t

Panel Data ↑ i – ind. t – time

$Y_{it} = X_{it}\beta + \mathcal{M}_{it}.$

consistent $\hat{\beta}$

① $E(\mathcal{M}_{it}\ X_i) = 0 \Rightarrow RE \Rightarrow$ OLS is fine

$Y_i = x_i\beta + \varepsilon_i \ldots$

homoskedasticity $E(\varepsilon_i^2) = 6^2 \longrightarrow$ OLS : BLUE

$Var(\hat{\beta}) = (XX)^{-1}6^2$

② $E(U_{it}\ U_{it}) =$

↳ cluster

→ FE → fixed effect

$E(X_{it}, \mathcal{M}_{it}) \neq 0. \Rightarrow$ OLS is not fine inconsistent

✓ xtreg Y x, fe cluster ( )

$Y_{it} - \overline{Y}_i = (X_{it} - \overline{X}_i), \beta + (\varepsilon_{it} - \overline{\varepsilon}_i)$

October 23rd.

Today.

1. reshape.
2. append / merge
3. fixed effect $\stackrel{\text{numerically}}{=}$ OLS putting in a bunch of fixed effect dummy

reshape long wage exp i(id) j(year)
wide i(id) j(year)

combine datasets

dataset 1

id race

1

2

$\xleftrightarrow{\text{merge}}$

dataset 2

id gender wage2001 wage2002

merge 1:1 id using dataset 2

after reshape long
merge 1:m id using dataset3

append.

dataset

id var1 var2

1

⋮

N_1

append using dataset 2

dataset

id var1 var2

N_1+1

⋮

N_2

OLS

* partitioned regression

$Y_i = X_i\beta + \varepsilon_i$

$X_i = [\underbrace{x_{1i}, x_{2i},}_{X_{1i}} \mid \underbrace{\dots x_{ki}}_{X_{2i}}]$

$y_i = X_{1i}\beta_1 + X_{2i}\underline{\underline{\beta_2}} + \varepsilon_i$

$\boxed{1}$ reg Y on X

$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \begin{matrix} \} \beta_1 \\ \} \beta_2 \end{matrix}$

$\boxed{2}$ ① reg Y_i on $X_{1i} \Rightarrow$ residual $\tilde{y}_i$

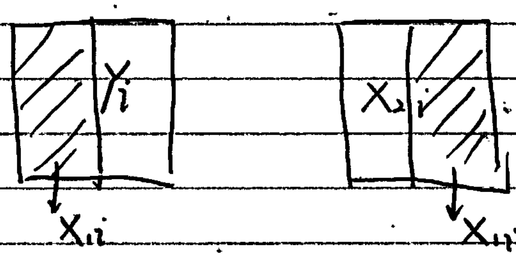
② reg X_{2i} on $X_{1i} \Rightarrow \tilde{X}_{2i}$

③ reg $\tilde{Y}_i$ on $\tilde{X}_{2i}$

① $Y_i = X_{1i}\gamma_1 + \eta_{1i} \Rightarrow \hat{\eta}_{1i} \Rightarrow \tilde{y}_i = y_i - X_{1i}\hat{\gamma}_1$

② $X_{2i} = X_{1i} \cdot \gamma_2 + \eta_{2i} \Rightarrow \hat{\gamma}_2 \Rightarrow \tilde{X}_{2i} = X_{2i} - X_{1i}\hat{\gamma}_2$

③ $\hat{\gamma}_3 \overset{\text{numerically}}{=} \hat{\beta}_2$



panel $Y_{it} = X_{it}\beta + \mu_{it}$

$\mu_{it} = \theta_i + \varepsilon_{it}$

xtreg y x, fe

$\Updownarrow$

OLS: $Y_{it} - \bar{Y}_i = (X_{it} - \bar{X}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$

i	t	Y	X	dumy1	dumy2 ...	dumy N
1	2001			1	0	
	2002			1		
	2003			1		
2	2001			0	1	0
	2002			0	1	
	2003			0	1	

reg y on x dumy *

① reg y on dumy * $\Rightarrow \tilde{y}$

② reg x on dumy * $\Rightarrow \tilde{X}$

③ reg $\tilde{y}$ on $\tilde{X}$

detour. $y_i = \beta + \varepsilon_i \Rightarrow \hat{\beta} = \bar{Y}_i$

$$\begin{bmatrix} y_{it} \end{bmatrix} = \begin{bmatrix} 1 & & \\ 1 & \vdots & \\ & & \vdots \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_N \end{bmatrix}$$

$$\gamma_1 = \frac{Y_{1,2001} + Y_{1,2002} + Y_{1,2003}}{3} = \bar{Y}_1$$

$$\tilde{Y}_1 = (Y_{it} - \bar{Y}_i)$$

$$Y_{it} - \bar{Y}_i = (X_{it} - \bar{X}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

check partitioned reg $\Leftarrow$

FE

2×1

Problem 1

$$E(Y_i | X_i) = \beta_0 e^{\beta_1 X_i}$$

$$E(Y_i - \beta_0 e^{\beta_1 X_i} | X_i) = 0$$

$$\Rightarrow E\left(f(X_i) \; (Y_i - \beta_0 e^{\beta_1 X_i})\right) = 0$$

$\boxed{1}$ $f(X_i) = \begin{bmatrix} 1 \\ X_i \end{bmatrix}$, [

$$E\left(\begin{bmatrix} 1 \\ X_i \end{bmatrix} (Y_i - \beta_0 e^{\beta_1 X_i}) \right) = 0$$

$$E\left(g(X_{i0}, \theta_0)\right) = 0$$ just identified. sample analog

$$\frac{1}{N} \sum_i g(x_i, \hat{\theta}) = 0$$

$$\begin{pmatrix} \frac{1}{N} \sum_i (Y_i - \hat{\beta}_0 e^{\hat{\beta}_1 X_i}) = 0 \\ \frac{1}{N} \sum_i (X_i Y_i - \hat{\beta}_0 X_i e^{\hat{\beta}_1 X_i}) = 0 \end{pmatrix}$$

$\boxed{2}$ $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, G^{-1} \Psi G^{-1})$

$G \to l \times k$

dim of moment condition — dim of parameter

$$\Psi = E[g(X_i, \theta_0) \, g(X_i, \theta_0)'] \quad \begin{matrix} l \times l \\ 2 \times 2 \end{matrix}$$

$\boxed{3}$ $\left.\frac{\partial E(Y_i | X_i)}{\partial X_i}\right|_{X_i = 0} = \beta_0 \beta_1 e^{\beta_1 X_i} = \beta_0 \beta_1$

$\downarrow$ $f(\hat{\theta}) \leftarrow \sqrt{n}(\hat{\theta} - \theta_0) \sim N(0, V)$

$$f(\hat{\theta}) = f(\theta_0) + \frac{\partial f(\tilde{\theta})}{\partial \theta'} (\hat{\theta} - \theta_0)$$

$$\sqrt{n}\left(f(\hat{\theta}) - f(\theta_0)\right) \xrightarrow{} N\left(0, \underbrace{\frac{\partial f(\theta_0)}{\partial \theta'}}_{k \times l} V \frac{\partial f(\theta_0)}{\partial \theta}\right)$$