Lecture 1:
Linear Model and OLS in Cross Section

Jesper Riis–Vestergaard Sørensen

University of Copenhagen, Department of Economics

Plan for Panel Data Lectures

Lecture 1: Linear model + OLS in cross section (W.4)

Lecture 2: Fixed effects + First differences (W.10)

Lecture 3: Random effects + Hausman test (W.10)

Lecture 4: Predetermined variables (W.11)

Cross Section Data

i	y	x^1	x^2	x^3	$\cdots$	x^K
1	y_1	1	x_1^2	x_1^3	$\cdots$	x_1^K
2	y_2	1	x_2^2	x_2^3	$\cdots$	x_2^K
3	y_3	1	x_3^2	x_3^3	$\cdots$	x_3^K
4	y_4	1	x_4^2	x_4^3	$\cdots$	x_4^K
5	y_5	1	x_5^2	x_5^3	$\cdots$	x_5^K
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	y_N	1	x_N^2	x_N^3	$\cdots$	x_N^K

Sampling Scheme for (Micro) Data

In this course **we will assume that:**

- Cross-sectional units (i) are *i*ndependent.
- Observations *id*entically distributed.

iid data

Focus is on asymptotics as number of cross section units grows without bound

- Finite-sample results rarely available.
- Limits ("arrows") understood as $N \to \infty$.
- Implicit assumption for asymptotics to be relevant:

N is large.

Today

- Another look at linear model and OLS in cross section.
- In part, refresher.
- ... but also more formal approach.
  - Will argue using asymptotics.
- Useful benchmark for later (panel) lectures.
  - Will essentially transform to cross section.

Outline

Model and Identification

Estimation and Consistency

Asymptotic Normality

Variance Estimation

Testing Linear Hypotheses

Model and Identification

Identification

$1 \times K$

Equation of interest in error form

$$y = \mathbf{x}\boldsymbol{\beta} + u$$

Assuming $E(u|\mathbf{x}) = 0$, we get

$$\boxed{E(y|\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}}$$

Identifying assumptions

OLS.1: $E(\mathbf{x}'u) = \mathbf{0}$

$K \times K$

OLS.2: rank $\boxed{E(\mathbf{x}'\mathbf{x})} = K$

Note: $E(u|\mathbf{x}) = 0$ is stronger than OLS.1

...but not necessary for identification of $\boldsymbol{\beta}$.

Identification

$x'y = x'x\beta + x'u$

Premultiply y equation by $\mathbf{x}'$ and take expectations

$$E(\mathbf{x}'y) = E(\mathbf{x}'\mathbf{x})\beta + \underbrace{E(\mathbf{x}'u)}_{=0}.$$

Under OLS.1.we can write

$$\boxed{E(\mathbf{x}'y) = E(\mathbf{x}'\mathbf{x})\beta}$$

Under OLS.2, $\boxed{E(\mathbf{x}'\mathbf{x})}$ is invertible, so

$$\boxed{\beta = [E(\mathbf{x}'\mathbf{x})]^{-1}E(\mathbf{x}'y)}$$

$E(\mathbf{x}'\mathbf{x})$ and $E(\mathbf{x}'y)$ features of joint distribution of y and $\mathbf{x}$

$\Longrightarrow$ **β is identified**

Estimation and Consistency

Estimation

Suppose we have cross section data

$\{(y_i, x_i)\}_{i=1}^{N}$

$$y_i = \mathbf{x}_i \beta + u_i, \quad i = 1, 2, \ldots, N.$$

Analogy principle: Replace unknowns with (consistent) estimators.

$\beta = (E[x'x])^{-1} E[x'y]$

Identification result + law of large numbers (LLN) suggest

$$\boxed{\hat{\beta} := \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i' y_i \right),} \qquad \text{(OLS)}$$
$$= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y},$$

$1 \times K$

- $\mathbf{X}$ $(N \times K)$ and $\mathbf{y}$ $(N \times 1)$ stack the $\mathbf{x}_i$s and y_is, respectively.

Consistency

$y_i = x_i\beta + u_i$

Inserting model,

unknown but random

$$\boxed{\hat{\beta} = \beta + \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'u_i\right).}$$

By random sampling + LLN,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'u_i \xrightarrow{p} E(\mathbf{x}'u) \overset{\text{OLS.1}}{=} \underbrace{\mathbf{0}}_{K\times 1}.$$

Similarly $N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i \rightarrow_p \underbrace{\boxed{E(\mathbf{x}'\mathbf{x})}}_{K\times K}$ which is invertible (OLS.2).

$\xrightarrow{p}$ element by element

Consistency

Applying Slutsky's theorem (W. Lemma 3.4),

$$\underbrace{\left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)}_{\downarrow p \;\; E(x'x)}^{-1} \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i' u_i\right)}_{\downarrow p \;\; E[x'u]} \xrightarrow{p} \underbrace{\left[E\left(\mathbf{x}'\mathbf{x}\right)\right]^{-1}}_{\text{exists}} \overbrace{E\left(\mathbf{x}'u\right)}^{= \underset{K\times 1}{0}}$$

$$= \left[E\left(\mathbf{x}'\mathbf{x}\right)\right]^{-1} \mathbf{0}$$

$$= \underset{K\times 1}{\mathbf{0}}.$$

Conclude: $\hat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}$.

OLS.1+OLS.2 imply consistency of OLS.

Inference?

Asymptotic Normality

Convergence in Distribution

Definition: A sequence $\{\mathbf{X}_N\}_1^\infty$ of G-dimensional random variables (r.v.'s) converges in distribution to the continuous r.v. $\mathbf{X}$ if for each $\mathbf{x} \in \mathbf{R}^G$,

$$\boxed{F_N(\mathbf{x}) \to F(\mathbf{x}).}$$

- Here $F_N(\cdot) := \mathrm{P}(\mathbf{X}_N \leqslant \cdot)$ denotes the cumulative distribution function (CDF) of $\mathbf{X}_N$
- ... and $F(\cdot) := \mathrm{P}(\mathbf{X} \leqslant \cdot)$ is the (continuous) CDF of $\mathbf{X}$.
- We write $\boxed{\mathbf{X}_N \to_d \mathbf{X}.}$ $\to_d$ / $\xrightarrow{d}$ convergence in distribution
- W. also uses $\sim_a$ $\underbrace{\text{("asymptotically distributed as")}}$.

Central Limit Theorem

- We will establish $\rightarrow_d$ by means of a central limit theorem.

Theorem (**CLT, W. Theorem 3.2**)

If $\{\mathbf{w}_i\}_1^\infty$ are iid G-dimensional r.v.'s with zero mean (+ finite variance), then

$$\sqrt{N}\frac{1}{N}\sum_i w_i = \boxed{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{w}_i} \xrightarrow{d} \mathrm{N}\left(\underset{G\times 1}{\mathbf{0}}, \overset{G\times G}{E\left(\mathbf{w}\mathbf{w}'\right)}\right).$$

- Hence, even if distribution of $\mathbf{w}_i$s unknown
- ... scaled average approximately normal *in large samples.*

Asymptotic Normality in OLS

Rewrite

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'u_i\right)$$

to get

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}}\boxed{\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{x}_i'u_i\right)}.$$

$E[\;] = 0$

OLS.2 implies $\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1} \to_p [E(\mathbf{x}'\mathbf{x})]^{-1}$.

By random sampling + CLT + OLS.1,

$E[(x'u)(x'u)']$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{x}_i'u_i \xrightarrow{d} \mathrm{N}\left(\underset{K\times 1}{\underline{\underline{0}}}, E\left(u^2\mathbf{x}'\mathbf{x}\right)\right).$$

$= E[u^2 x'x]$

Combining Modes of Convergence

Product Rule: If

1. $\mathbf{Y}_N \to_p \mathbf{C}$ constant (matrix), and
2. $\mathbf{Z}_N \to_d \mathbf{Z}$,

then $\boxed{\mathbf{Y}_N \mathbf{Z}_N \to_d \mathbf{CZ}.}$

Implicit: Y_Ns & Z_Ns are conformable

Warning: Constancy of $\mathbf{C}$ cannot be disposed of.

- Will use this product rule again and again...

Asymptotic Normality

constant matrix

It follows that

$$\sqrt{N}(\hat{\beta}-\beta) = \overbrace{\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}}^{\xrightarrow{p}[E(x'x)]^{-1}} \overbrace{\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{x}_i'u_i\right)}^{\xrightarrow{d} N(\underline{0},\, E(u^2x'x))}$$

$$\xrightarrow{d} \underbrace{[E(\mathbf{x}'\mathbf{x})]^{-1}}_{\text{“}C\text{”}} \underbrace{\mathrm{N}\left(\mathbf{0}, E\left(u^2\mathbf{x}'\mathbf{x}\right)\right)}_{\text{“}Z\text{”}}. \quad \text{(Product Rule)}$$

Normal property:

1. If $\mathbf{C}$ is constant (matrix)
2. and $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{V})$,

then $\mathbf{CZ} \sim \mathrm{N}(\mathbf{0}, \mathbf{CVC}')$.

(Normal family closed under affine transformations)

Asymptotic Normality

$\sqrt{N}$ (OLS - parameters)

Conclude:

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \overbrace{\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}}^{\text{sandwich}}\right),$$

$$\text{where} \quad \mathbf{A} := \boxed{E\left(\mathbf{x}'\mathbf{x}\right)},$$

$$\mathbf{B} := E\left(u^2\mathbf{x}'\mathbf{x}\right),$$

"meat"

Under OLS.1+OLS.2, OLS is ($\sqrt{N}$-)asymptotically normal.

Variance Estimation

Asymptotic Variance Estimation

$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d}$

treat as $\overset{d}{=} / \sim$

Asymptotic distribution suggests approximation

$$\boxed{\hat{\boldsymbol{\beta}} \overset{d}{\approx} \mathrm{N}\left(\boldsymbol{\beta}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}/N\right).}$$

(Treat "$\rightarrow_d$" as "$\sim$".)

"$\overset{d}{\approx}$" reads as "approximately distributed as."

Should be good approximation in "large" samples.

Asymptotic Variance Estimation

$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(0, A^{-1} B A^{-1})$

Potential source of confusion:

Limit theory implies

$$\operatorname{var}[\sqrt{N}(\hat{\beta} - \beta)] \to \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} =: \mathbf{V}.$$

So $\mathbf{V}$ is the asymptotic variance of $\sqrt{N}(\hat{\beta} - \beta)$.

Call $\boxed{\operatorname{Avar}(\widehat{\beta}) := \mathbf{V}/N}$ the asymptotic variance of $\widehat{\beta}$.

"$\widehat{\mathbf{V}}/N$ consistently estimates $\operatorname{Avar}(\widehat{\beta})$" means "$\widehat{\mathbf{V}} \to_p \mathbf{V}$."

Convenient—but imprecise—shorthand.

Asymptotic Variance Estimation

$= \hat{A}^{-1}\hat{B}\hat{A}^{-1}/N$

A consistent estimator of $\mathrm{Avar}(\widehat{\beta})$ is

$= \widehat{\mathrm{Avar}(\beta)}$

$$\boxed{(\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{x}_i'\mathbf{x}_i\right)(\mathbf{X}'\mathbf{X})^{-1}}, \quad \boxed{\widehat{u}_i := y_i - \mathbf{x}_i\widehat{\beta}.}$$

Just

$$\boxed{\widehat{\mathbf{V}}/N := \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}/N}$$

with

$$\boxed{\widehat{\mathbf{A}} := \frac{1}{N}\mathbf{X}'\mathbf{X} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i} \quad \text{and} \quad \boxed{\widehat{\mathbf{B}} := \frac{1}{N}\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{x}_i'\mathbf{x}_i.}$$

Robust to heteroskedasticity.

1st $E[\cdot] \to \frac{1}{N}\sum_i$

$\frac{1}{N}\sum_i u_i^2 x_i'x_i$

Efficiency

constant scalar

$\boxed{\textbf{OLS.3}: \; E(u^2 \mathbf{x}'\mathbf{x}) = \sigma^2 E(\mathbf{x}'\mathbf{x})}$ $\Leftrightarrow \; B = \sigma^2 A$

- Implies OLS also asymptotically efficient
- Simpler variance estimator

$V = A^{-1} B A^{-1} / N$

$ = \sigma^2 A^{-1} / N$

$$\widehat{\text{Avar}(\hat{\beta})} = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}, \quad \boxed{\hat{\sigma}^2 := \frac{1}{N-K} \sum_{i=1}^{N} \hat{u}_i^2.}$$

degrees of freedom correction

If OLS 3. is violated?

$\hat{\beta}$ not necessarily efficient.

Inconsistent $\widehat{\text{Avar}(\hat{\beta})}$ $\Rightarrow$ rely on robust version for inference.

Testing Linear Hypotheses

Testing Linear Hypotheses

Interest in $\mathrm{H}_0 : \mathbf{R}\beta = \mathbf{r}$.   null hypothesis

- $\mathbf{R}$: $Q \times K$ with rank $\mathbf{R} = Q \leqslant K$
- $\mathbf{r}$: $Q \times 1$

$\hat{\beta} \neq \beta$

$\mathbf{R}\hat{\beta} \approx \mathbf{R}\beta \overset{H_0}{=} \mathbf{r}$

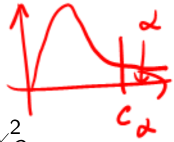
Wald statistic:

$$W := \overbrace{(\mathbf{R}\widehat{\beta} - \mathbf{r})' \left[\mathbf{R}\widehat{\mathrm{Avar}(\widehat{\beta})}\mathbf{R}'\right]^{-1} (\mathbf{R}\widehat{\beta} - \mathbf{r})}^{\text{quadratic}}.$$

$\underbrace{\mathbf{R}\hat{\beta} - \mathbf{r}}_{\text{small}} \approx 0$

Under H_0, $W \to_d \chi^2_Q$. $\leftarrow$ = # restrictions

Wald test:



Reject H_0 at level $\underset{\in(0,1)}{\alpha} \Leftrightarrow W >$ $(1-\alpha)$-quantile of χ^2_Q.

Robust to heteroskedasticity if robust $\widehat{\mathbf{V}}$ used.

On Wald Statistic Form

Fact: If $\mathbf{Z} \sim \mathrm{N}\left(\mathbf{0}_{G\times 1}, \boldsymbol{\Sigma}\right)$ then $\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} \sim \chi^2_G$.

- Multivariate version of $Z \sim \mathrm{N}\left(0, \sigma^2\right) \Rightarrow \left(Z/\sigma\right)^2 \sim \chi^2_1$.

OLS.1–2 imply $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \approx_d \mathrm{N}\left(\mathbf{0}_{K\times 1}, \mathbf{V}/N\right)$.

Linearly transform: $\underbrace{\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}}_{\text{“}\mathbf{Z}\text{”}} \approx_d \mathrm{N}\left(\mathbf{0}_{Q\times 1}, \underbrace{\mathbf{R}\left(\mathbf{V}/N\right)\mathbf{R}'}_{\text{“}\Sigma\text{”}}\right)$

Conclude:

$$\overbrace{(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})'}\left[\mathbf{R}\underbrace{\left(\mathbf{V}/N\right)}\mathbf{R}'\right]^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}) \overset{d}{\approx} \chi^2_Q.$$

Under H_0, $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.

Wald arises from consistent $\mathrm{Avar}(\widehat{\boldsymbol{\beta}}) = \mathbf{V}/N$ estimator.

Example Hypotheses ($K = 3$)

1. "$\beta_1 = \beta_2$" corresponds to

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \quad \text{and} \quad r = 0.$$

$\beta_1 - \beta_2 = 0$

2. "$\beta_1 + \beta_2 = 1$" corresponds to

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad r = 1.$$

3. "$\beta_1 = \beta_2 = \beta_3$" corresponds to

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$\beta_1 - \beta_2 = 0$
$\beta_2 - \beta_3 = 0$

Q: what about $\beta_1 = \beta_3$?

4. "$\beta_1 = \beta_2 = \beta_3 = 0$" corresponds to

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

I_3