



Linear Models for Panel Data

Advanced Microeconometrics

Anders Munk-Nielsen

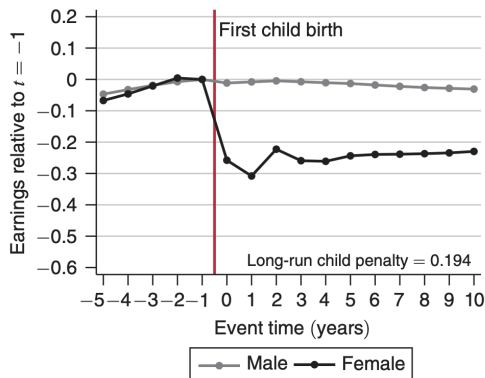
2022



Introduction

Child Penalty

Panel A. Earnings



Source: Kleven, Landais & Sørensen (2019; AEJ Applied Economics).

They use Fixed Effects, whereas previous literature used Pooled OLS.

The female child penalty in wages

We want to know if females have a higher wage-penalty and estimate the model

$$\text{wage}_{it} = \beta_1 \mathbf{1}\{\text{child}\}_{it} + \beta_2 \mathbf{1}\{\text{child}\}_{it} \mathbf{1}\{\text{female}\}_i + c_i + u_{it}.$$

where $\mathbf{1}\{\text{child}\}$ is a dummy for having a child.

Discuss

1. Why is there no separate gender dummy? ($\beta_3 \mathbf{1}\{\text{female}\}$) this would be wages for females w/out children
2. Based on the graph, what are β_1 and β_2 approximate? we are interested in variation of wages within ind that have children
Interpret this. \beta_1 is around zero - having a child does not seem to impact males
\beta_2 is approx -0.25 - having a child penalizes earnings by about .25 percent
3. Give *intuitive* (meaning: real-world) explanations for correlation (positive or negative) between c_i and not those do not have children
 - 3.1 $\mathbf{1}\{\text{child}\}$, E(\beta_{1c_i}) >= 0 -> men choose jobs with little-to-no paternity leave(?) / preferring work to child-care
 - 3.2 $\mathbf{1}\{\text{child}\} \mathbf{1}\{\text{female}\}$, E(\beta_{2c_i}) < 0 -> women choosing jobs with lower wages after child-birth / preferring child-care to work

Positive view

Panel B. Child-related gender inequality versus education-related gender inequality (post-child effects versus pre-child effects)

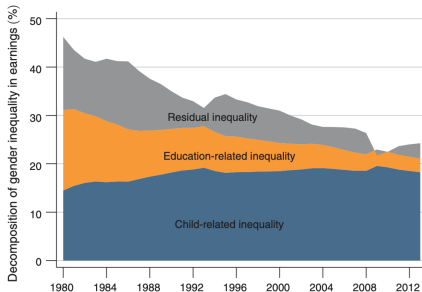


FIGURE 5. DECOMPOSING GENDER INEQUALITY IN EARNINGS

Over time: total inequality ↓, so now almost only child-related inequality remains.

Where are we in the course?

Part	Topic	Parameterization non-linear	Estimation non-linear	Dimension $\dim(x)$	Numerical optimization	M-estimation (Part III)	Outcome (y_i)	Panel (c_i)
I	OLS	÷	÷	low	÷	✓	\mathbb{R}	✓
II	LASSO	÷	✓	high	✓	÷	\mathbb{R}	÷
	Probit	✓	✓	low	✓	✓	$\{0, 1\}$	÷
	Tobit	✓	✓	low	✓	✓	$[0; \infty)$	÷
IV	Logit	✓	✓	low	✓	✓	$\{1, 2, \dots, J\}$	÷
	Sample selection	✓	✓	low	✓	✓	\mathbb{R} and $\{0, 1\}$	÷
	Simulated Likelihood	✓	✓	low	✓	✓	Any	✓
	Quantile Regression	÷	✓	(low)	✓	✓	\mathbb{R}	÷
	Non-parametric	✓	(✓)	∞	÷	÷	\mathbb{R}	÷

Panel model

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}.$$

Assumptions	Rank	Description	Estimator				
			POLS	BE	RE	FE	FD
$E(u_{it} \mathbf{x}_{it}, c_i) = 0, c_i = c,$	$\mathbf{X}'\mathbf{X}$	No individual effects	✓+★	✓	✓	✓	✓
$E(u_{it} \mathbf{X}_i, c_i) = 0, c_i \sim \text{IID}(0, \sigma_c^2),$	$\mathbf{X}'\mathbf{X}$	Random effects	✓	✓	✓+★	✓	✓
$E(u_{it} \mathbf{X}_i, c_i) = 0, E(c_i \mathbf{x}_{it}) \neq 0,$	$\check{\mathbf{X}}'\check{\mathbf{X}}$	Fixed effects	÷	÷	÷	✓	✓
$E(u_{it} \mathbf{X}_i, c_i) = 0, E(c_i \mathbf{x}_{it}) \neq 0, \text{ and } u_{it} \text{ IID}$	$\check{\mathbf{X}}'\check{\mathbf{X}}$	Fixed effects	÷	÷	÷	✓+★	✓
$E(u_{it} \mathbf{X}_i, c_i) = 0, E(c_i \mathbf{x}_{it}) \neq 0 \text{ and } \Delta u_{it} \text{ IID}$	$\Delta\mathbf{X}'\Delta\mathbf{X}$	Fixed effects	÷	÷	÷	✓	✓+★
$E(u_{it} \mathbf{x}_{it}, c_i) \neq 0$	–	Endogeneity	÷	÷	÷	÷	÷

(÷ = inconsistent, ✓ = consistent, ★ = efficient)

- **Average:** a “bar” denotes an average,

$$\bar{\mathbf{x}}_i \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{x}_{it}$$

- **Transformations:**

- Demeaning:

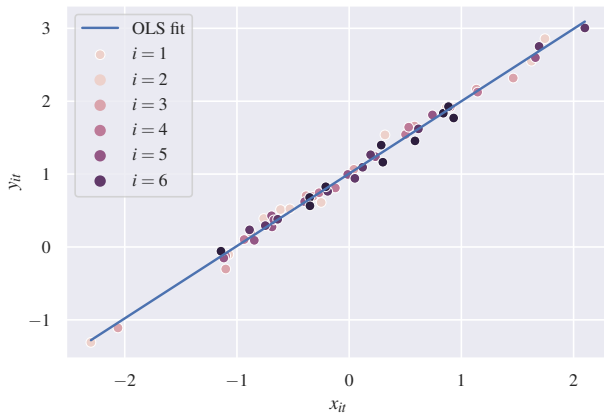
$$\ddot{\mathbf{x}}_{it} \equiv \mathbf{x}_{it} - \bar{\mathbf{x}}_i,$$

- First-differences:

$$\Delta \mathbf{x}_{it} \equiv \mathbf{x}_{it} - \mathbf{x}_{it-1}.$$

- **ALWAYS** define explicitly what you mean by mathematical symbols.
- **Estimator:** A “hat” denotes an estimator, e.g. $\hat{\beta}$ is an estimate of β .

No Unobserved Effects: $c_i = 0$



Zooming out on y



Adding an IID c_i : $E(x'_i c_i) = 0$ (Random Effects)



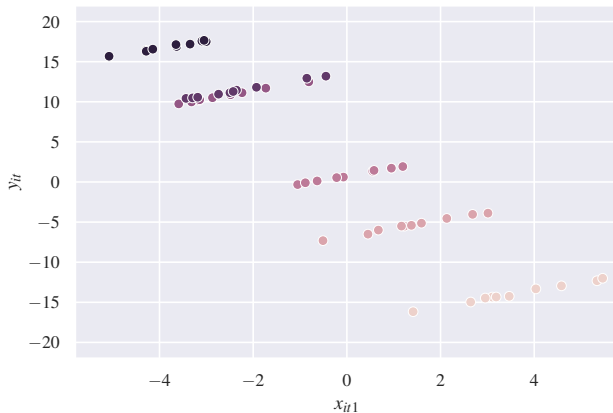
Adding an IID c_i : $E(\mathbf{x}'_i c_i) = 0$ – pooled OLS



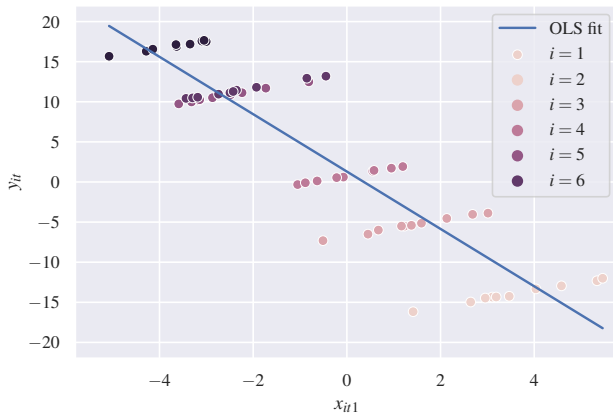
Zooming out on x



Allowing Endogeneity: $E(x'_{it}c_i) \neq 0$ (Fixed Effects)



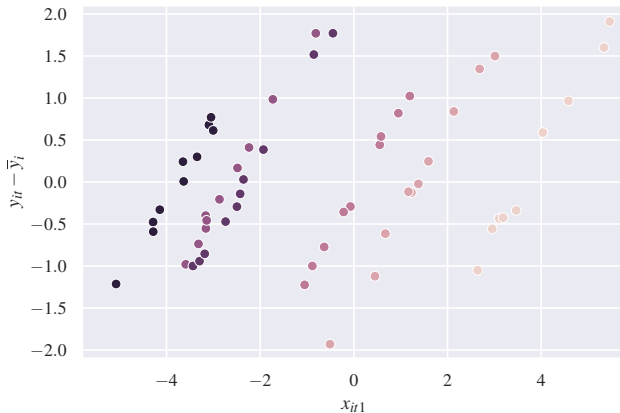
Pooled OLS is inconsistent



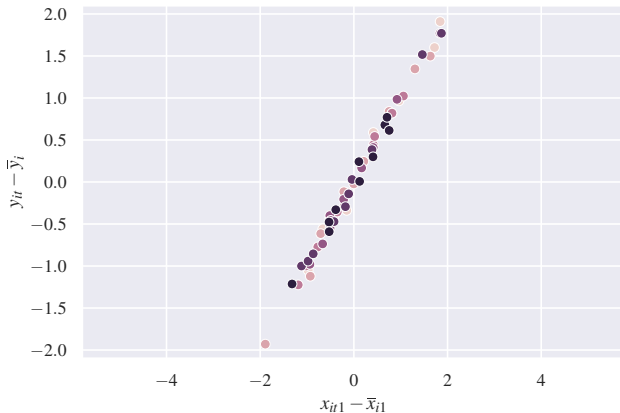
Demeaning y_{it}



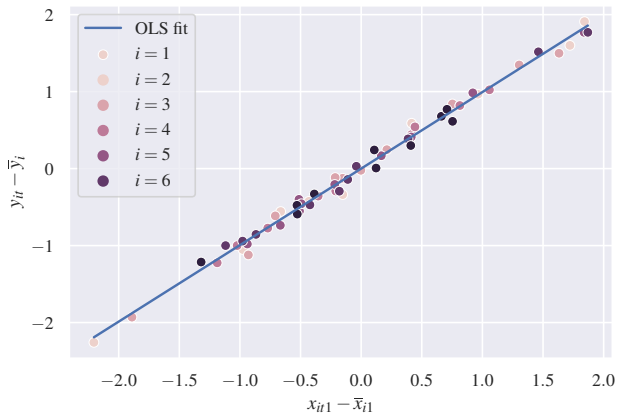
Demeaning y_{it} ... and zooming



Also demeaning x_{it}



POLS on Demeaned Data



Model

General Model

$$y_{it} = c_{it} + \mathbf{x}_{it}\beta_{it} + u_{it}.$$

- **Suppose** we assume $u_{it} \sim \text{IID}(0, \sigma_u^2)$
(i.e. u_{it} independent of all variables, so in particular $E(u_{it} | \mathbf{x}_{it}, c_{it}, \beta_{it}) = 0$)
 - **Discuss:** *"This model is too general and is not estimable ..."*
(CT, p. 699)
- **Moving forward:** our assumptions will impose structure.
Throughout, we at least assume
 - $\beta_{it} = \beta$ (common slope), rules out many interesting effects
 - $c_{it} = c_i$ (time-constant individual specific effect)

Linear panel data model

Model

$$y_{it} = c_i + \mathbf{x}_{it}\beta + u_{it}.$$

- **Error term:** it is fruitful to write the model in terms of a *composite error term*, denoted v_{it} .

- **Intercept?** We can either

1. Assume $E(c_i) = c$ and have no constant in \mathbf{x}_{it}

$$y_{it} = c + \mathbf{x}_{it}\beta + v_{it}, \quad v_{it} = c_i - c + u_{it}$$

2. Assume $E(c_i) = 0$ and have a constant in \mathbf{x}_{it} [preferred]

$$y_{it} = \mathbf{x}_{it}\beta + v_{it}, \quad v_{it} = c_i + u_{it}$$

- **Usefulness:** map assumptions about c_i, u_{it} into behavior of v_{it} and use results about cross-sectional OLS.

Assumption: strict (or strong) exogeneity (SE)

$$E(u_{it} | c_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}) = 0.$$

Assumption: contemporaneous exogeneity (CE)

$$E(u_{it} | c_i, \mathbf{x}_{it}) = 0.$$

- SE rules out y_{it-1} as a regressor in \mathbf{x}_{it} .
 - We will prove this later...
 - ... but basically, y_{it} is in \mathbf{x}_{it+1} ...
- **Distinction:** Conditional \Rightarrow unconditional exogeneity
 - Conditional: $E(u_{it} | \mathbf{x}_{it}) = 0$ stricter
 - Unconditional: $E(\mathbf{x}_{it} u_{it}) = 0$ weaker

What do we observe?

- **Model components:** $y_{it}, \mathbf{x}_{it}, c_i, u_{it}$. Only $(y_{it}, \mathbf{x}_{it})$ observed.
- **We can compute** $E(y_{it}|\mathbf{x}_{it})$ for different values of \mathbf{x}_{it}
 - That is, we can maybe hope to estimate

$$\frac{\partial E(y_{it}|\mathbf{x}_{it})}{\partial x_{itk}}.$$

- **The problem:** By strict exogeneity, the model implies

$$E(y_{it}|\mathbf{x}_{it}) = E(c_i|\mathbf{x}_{it}) + \mathbf{x}_{it}\beta.$$

- If $E(c_i|\mathbf{x}_{it})$ varies with \mathbf{x}_{it} , then that is reflected in $\frac{\partial E(y_{it}|\mathbf{x}_{it})}{\partial x_{itk}}$
- **Note:** the important part is whether $E(c_i|\mathbf{x}_{it}) = E(c_i)$ (not a function of \mathbf{x}_{it}),
 - ...not whether $E(c_i) = 0$
 - (if $E(c_i) = c \neq 0$, this cannot be separately identified from the intercept)

Proving Inconsistency

- **Proposition:** POLS is inconsistent if $E(\mathbf{x}'_{it}c_i) \neq \mathbf{0}$ or $E(\mathbf{x}'_{it}u_{it}) \neq \mathbf{0}$.
- **Proof:** By definition, POLS is

$$\begin{aligned}\hat{\beta}_{POLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{c} + \mathbf{u}) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{c} + \mathbf{X}'\mathbf{u}).\end{aligned}$$

- Take the plim and note that $\text{plim}(N^{-1}T^{-1}\mathbf{X}'\mathbf{X}) = E(\mathbf{x}'\mathbf{x})$,

$$\text{plim}\hat{\beta}_{POLS} = \beta + E(\mathbf{x}'\mathbf{x})^{-1}[E(\mathbf{x}'\mathbf{c}) + E(\mathbf{x}'\mathbf{u})]$$

by contemporaneous exogeneity.

- **Thus:** $\text{plim}\hat{\beta}_{POLS} = \beta$ if $E(\mathbf{x}\mathbf{c}) = \mathbf{0}$ and $E(\mathbf{x}\mathbf{u}) = \mathbf{0}$, but not otherwise. ■

- **Now:** different strategies based on $E(\mathbf{x}c_i)$.
- Under **exogeneity** of the individual effect, $E(\mathbf{x}c_i) = \mathbf{0}$:
 - Pooled OLS is consistent
 - Random Effects Estimator is efficient
- Under **endogeneity**, $E(\mathbf{x}c_i) \neq \mathbf{0}$,
 - Fixed Effects (FE) estimator
 - First Differences (FD) estimator
 - **Common feature:** Transform data to eliminate c_i
 - **Implication:** only within-individual time variation identifies β

Fixed Effects Model

Model

$$y_{it} = c_i + \mathbf{x}_{it}\beta + u_{it}.$$

- **Assume:**

- Strict exogeneity: $E(u_{it}|\mathbf{X}_i, c_i) = 0$ ($\mathbf{X}_i \equiv (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})$)
- Rank condition: (to be specified)

- **Idea:** perform *within transformation*,

$$\ddot{a}_{it} \equiv a_{it} - \bar{a}_i, \quad \bar{a}_i \equiv \frac{1}{T} \sum_{t=1}^T a_{it}$$

for variables $\ddot{y}_{it}, \ddot{x}_{it}, \ddot{u}_{it}$.

Within transformation

- **Starting point:** model equation

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}.$$

- **Transformed** equation

$$\begin{aligned} y_{it} - \bar{y}_i &= (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\beta + \overbrace{c_i - \bar{c}_i}^{=0} + u_{it} - \bar{u}_i \\ \check{y}_{it} &= \check{\mathbf{x}}_{it}\beta + \check{u}_{it}. \end{aligned}$$

- **Great success!** c_i is eliminated from the equation!
- **Note:** no assumptions regarding c_i – except
 - Time-constant,
 - Functional form (additively separable).
 - \Rightarrow we can allow for *arbitrary correlation between \mathbf{x} and c_i*

Fixed Effects (FE) Estimator

$$\hat{\beta}_{FE} = (\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'\ddot{\mathbf{y}}.$$

(also called the Within Estimator)

- **Does this work?** Same steps as cross-sectional OLS:
 - Focus on the equation $\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\beta + \ddot{\mathbf{u}}_{it} \dots$
 - ... check conditions for OLS on this:
- **Exogeneity:** $E(\ddot{u}_{it}\ddot{\mathbf{x}}_{it}) = \mathbf{0}$.
- **Rank condition:** $\ddot{\mathbf{X}}'\ddot{\mathbf{X}}$ must be invertible:

FE Assumptions for Consistency

1. $E(\ddot{\mathbf{x}}_{it} \ddot{u}_{it}) = \mathbf{0}$
2. $\ddot{\mathbf{X}}'\ddot{\mathbf{X}}$ must be invertible

Challenge: translate assumptions on model primitives $(\mathbf{x}_{it}, u_{it}, c_i)$ to assumptions on transformed data $(\ddot{\mathbf{x}}_{it}, \ddot{u}_{it})$

- **Exogeneity:** contemporaneous exogeneity ($E(u_{it}|\mathbf{x}_{it}) = 0$) is insufficient

$$E(\ddot{\mathbf{x}}_{it} \ddot{u}_{it}) = E(\mathbf{x}_{it} u_{it}) - T^{-1} \sum_{s=1}^T [E(\mathbf{x}_{is} u_{it}) + E(\mathbf{x}_{it} u_{is})] + E(\bar{\mathbf{x}}_i \bar{u}_i)$$

- **Strict exogeneity:** $E(u_{it}|\mathbf{x}_i) = 0 \Rightarrow E(u_{it}\mathbf{x}_{is}) = 0 \forall t, s$.
- **Alternative formulation** of strict exogeneity:

$$E(y_{it}|\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i).$$

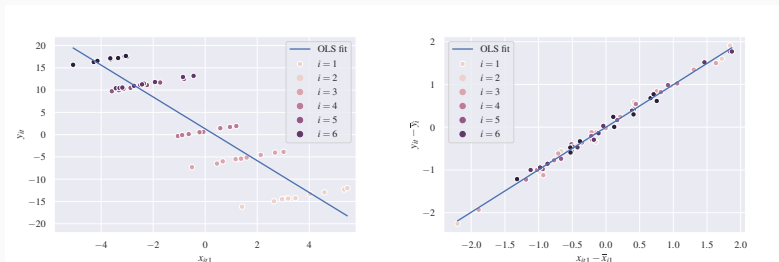
- E.g. violated for dynamic models (\mathbf{x}_{it-1} may be relevant)

FE Assumptions for Consistency

1. $E(\ddot{\mathbf{x}}_{it} \ddot{u}_{it}) = \mathbf{0}$
2. $\ddot{\mathbf{X}}'\ddot{\mathbf{X}}$ must be invertible

- **Rank condition:** what does this rule out?
 - **Answer:** _____ variables
 - **Examples:** birth year, sex, genes, ethnic origin, etc.
- **Discuss:** in our “effect of education on wages,” what sort of data requirements does this pose?

FE Graphically



- **Inconsistency** of POLS caused by $\text{Corr}(\mathbf{x}_{it}, c_i)$
- **Demeaning** \Rightarrow points are neatly “on a line”

Example from Problem Set 4

Fertility and Female Labor

Consider this model of female occupation

$$\mathbf{1}\{i \text{ works in year } t\} = \mathbf{x}_{it}\beta + c_i + u_{it}.$$

where \mathbf{x}_{it} contains $\mathbf{1}\{\text{gives birth in year } t\}$. Which of the following stories violate strict/contemporaneous exogeneity

- Alice experiences health problems related to her birth and has stop working.
- Beth foresees a pandemic next year and is worried she may get fired, so she gets pregnant to be legally protected.
- Cynthia's firm shuts down. She is bored and accidentally has a baby. Now she abhors the thought of job interviews.
- Some men hate pregnant women.

Implementation with matrix algebra

- **Cool result:** you can do demeaning using matrices!
 - \Rightarrow faster computation

Demeaning matrix

$$\mathbf{Q}_T \equiv \mathbf{I}_{T \times T} - \frac{1}{T} \mathbf{1}_{T \times T} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} - \begin{pmatrix} T^{-1} & \dots & T^{-1} \\ \vdots & \ddots & \vdots \\ T^{-1} & \dots & T^{-1} \end{pmatrix}$$

- Then $\ddot{\mathbf{x}}_i = \mathbf{Q}_T \mathbf{x}_i$.
- **The exercises** give hands on experience with this

Alternative Idea: LSDV

- **Alternative idea:** why not just estimate the c_i as parameters?
 - \Rightarrow **Least Squares Dummy Variables (LSDV)** Estimator
- **How:** simply add N dummies, $\{\mathbf{1}_i\}_{i=1}^N$, to \mathbf{x}_{it} .
- **Result:** numerically the same as FE.
- **Drawbacks:**
 - **Computationally:** $(\mathbf{X}'\mathbf{X})^{-1}$ takes a long time to compute ($N - 1 + K \times N - 1 + K$ inversion) and takes up a ton of RAM
 - **Theoretically:** Since $N \rightarrow \infty$ but $T \not\rightarrow \infty$, $\dim(\beta)$ increases in N !
 - High-dimensional problem; what does consistency mean? What is the “true” β , ∞ -dimensional?
 - Known as the “incidental parameters problem”
- **Conclusion:** we always prefer OLS on within-transformed data to LSDV!

Inference

Asymptotic Distribution

- **Writing out:**

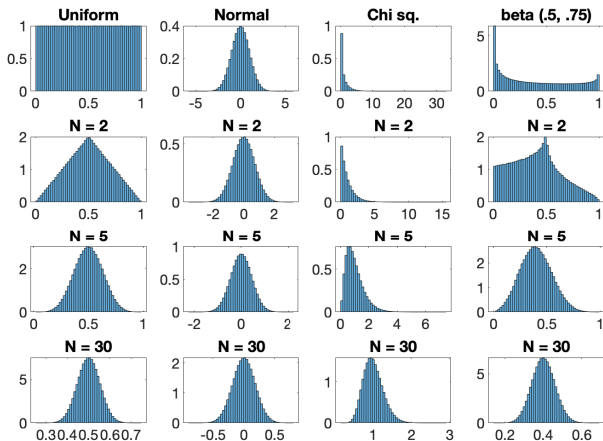
$$\begin{aligned}\hat{\beta}_{FE} &= (\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'(\ddot{\mathbf{X}}\beta + \ddot{\mathbf{u}}) \\ &= \beta + (\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'\ddot{\mathbf{u}}\end{aligned}$$

- **Similarly** to cross-sectional OLS, if our assumptions hold on the *transformed* data, we get

$$\sqrt{N}(\hat{\beta}_{FE} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}).$$

- The “bread” is similar: $\mathbf{A} \equiv E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i)$ ($\ddot{\mathbf{X}}_i$ is $T \times K$ so \mathbf{A} is $K \times K$).
- The “meat” is different
- **Normality** comes from the CLT again
 - “factor \times average \xrightarrow{d} Normal”
 - $\frac{1}{N} \sum_i \sum_t \ddot{\mathbf{x}}'_{it} \ddot{u}_{it}$ is our average,
 - $\frac{1}{N} \sum_i \sum_t \ddot{\mathbf{x}}'_{it} \ddot{\mathbf{x}}_{it}$ is the “factor.” (Note: $\frac{1}{T}$ cancelled out.)

Central Limit Theorem: Illustration



Columns are underlying distribution families, rows show the distribution of an *average* based on N draws. Note the x axis scale changing over rows!

Variance

- **Question:** what is the asymptotic variance?
- **CLT** says for a sample $\{\mathbf{w}_i\}_{i=1}^N$ with $E(\mathbf{w}_i) = \mathbf{0}_{1 \times K}$, then

$$\frac{1}{\sqrt{N}} \sum_i \mathbf{w}_i \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{B}), \quad \mathbf{B} \equiv V(\mathbf{w}_i).$$

- **Our case:** $\mathbf{w}_i := \sum_{t=1}^T \ddot{\mathbf{x}}'_{it} \ddot{u}_{it} \ (K \times 1)$
- **Variance of a sum:** recall $V(\sum_j \mathbf{a}_j) = \sum_j \sum_{j'} \text{Cov}(\mathbf{a}_j, \mathbf{a}_{j'})$.

$$\begin{aligned} V\left(\sum_{t=1}^T \ddot{\mathbf{x}}'_{it} \ddot{u}_{it}\right) &= \sum_{t=1}^T \sum_{s=1}^T \text{Cov}(\ddot{\mathbf{x}}'_{it} \ddot{u}_{it}, \ddot{\mathbf{x}}'_{is} \ddot{u}_{is}) \\ &= \sum_{t=1}^T \sum_{s=1}^T E\left[\ddot{\mathbf{x}}'_{it} \ddot{u}_{it} (\ddot{\mathbf{x}}'_{is} \ddot{u}_{is})'\right] \\ &= \sum_{t=1}^T \sum_{s=1}^T E(\ddot{\mathbf{x}}'_{it} \ddot{u}_{it} \ddot{u}_{is} \ddot{\mathbf{x}}_{is}). \end{aligned}$$

Since $\text{Cov}(\mathbf{a}, \mathbf{b}) = E(\mathbf{a}\mathbf{b}') - E(\mathbf{a})E(\mathbf{b})'$, $E(\ddot{u}_{it}) = 0$, and $(\mathbf{a}\mathbf{b})' = \mathbf{b}'\mathbf{a}'$.

- We showed:

$$V\left(\sum_{t=1}^T \ddot{\mathbf{x}}'_{it} \ddot{u}_{it}\right) = \sum_{t=1}^T \sum_{s=1}^T E(\ddot{\mathbf{x}}'_{it} \ddot{u}_{it} \ddot{u}_{is} \ddot{\mathbf{x}}_{is}) \quad (K \times K)$$

- Law of Iterated Expectations (LIE) gives

$$E(\ddot{\mathbf{x}}'_{it} \ddot{u}_{it} \ddot{u}_{is} \ddot{\mathbf{x}}_{is}) = E[\ddot{\mathbf{x}}'_{it} E(\ddot{u}_{it} \ddot{u}_{is} | \ddot{\mathbf{X}}_i) \ddot{\mathbf{x}}_{is}]$$

- In matrix form with $\ddot{\mathbf{X}}_i$ as $T \times K$:

$$V(\ddot{\mathbf{X}}_i \ddot{\mathbf{u}}_i) = E\left[\underbrace{\ddot{\mathbf{X}}'_i E(\ddot{\mathbf{u}}_i \ddot{\mathbf{u}}'_i | \ddot{\mathbf{X}}_i) \ddot{\mathbf{X}}_i}_{T \times T}\right] \quad (K \times K)$$

Assumption FE.3 (Homoskedasticity)

$$E(\mathbf{u}_i \mathbf{u}'_i | \mathbf{x}_i, c_i) = \sigma_u^2 \mathbf{I}_T.$$

Adding **FE.1**, this implies $V(\mathbf{u}_i | \mathbf{x}_i, c_i) = \sigma_u^2 \mathbf{I}_T$.

Asymptotically, under FE.1 and FE.2:

$$\sqrt{N} \left(\hat{\beta}_{FE} - \beta \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}),$$

$$\mathbf{A} \equiv E(\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i), \quad \mathbf{B} \equiv E \left[\ddot{\mathbf{X}}_i' E(\ddot{\mathbf{u}}_i \ddot{\mathbf{u}}_i' | \ddot{\mathbf{X}}_i) \ddot{\mathbf{X}}_i \right].$$

Panel-robust Variance Estimator

$$\widehat{\text{Avar}}(\hat{\beta}_{FE}) = (\ddot{\mathbf{X}}' \ddot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \ddot{\mathbf{X}}_i' \hat{\ddot{\mathbf{u}}}_i \hat{\ddot{\mathbf{u}}}_i' \ddot{\mathbf{X}}_i \right) (\ddot{\mathbf{X}}' \ddot{\mathbf{X}})^{-1},$$

where $\hat{\ddot{u}}_{it} = \ddot{y}_{it} - \ddot{\mathbf{x}}_{it} \beta_{FE}$.

Variance Estimator under Homoskedasticity (FE.3)

$$\begin{aligned} \widehat{\text{Avar}}(\hat{\beta}_{FE}) &= \hat{\sigma}_u^2 (\ddot{\mathbf{X}}' \ddot{\mathbf{X}})^{-1}, \\ \hat{\sigma}_u^2 &= \frac{\sum_{i=1}^N \sum_{t=1}^T \hat{\ddot{u}}_{it}^2}{NT - N - K}. \end{aligned}$$

- **Under FE.3:** We estimate

$$\hat{\sigma}_u^2 = \frac{\sum_i \sum_t \hat{u}_{it}^2}{NT - N - K}.$$

- **Usual df. correction:** “ $-K$ ”: for efficiency, no effect asymptotically (as $N \rightarrow \infty$)
 - We have estimated K parameters in $\hat{\beta}_{FE}$
- **New correction:** “ $-N$ ”: This *does* matter – the estimator is *inconsistent* without it.
 - We have estimated N “means”
 - LSDV: Here we explicitly estimated N dummies
 - ... and the two estimators are identical.

Digression: Cluster-robust Variance Estimation?

- We assumed IID across i (but not over time within i)
- In other settings, observations may be grouped in other forms of “clusters”.
 - E.g. classrooms, markets, regions, etc.
- First note that we can always write

$$\begin{aligned}V(\mathbf{X}'\mathbf{u}) &= V\left(\sum_{t=1}^T \sum_{i=1}^N \mathbf{x}'_{it} u_{it}\right) \\&= \sum_{t=1}^T \sum_{i=1}^N \sum_{s=1}^T \sum_{j=1}^N \text{Cov}(\mathbf{x}'_{it} u_{it}, \mathbf{x}'_{js} u_{js}) \\&= \sum_{t=1}^T \sum_{i=1}^N \sum_{s=1}^T \sum_{j=1}^N E\left[\mathbf{x}'_{it} E(u_{it} u_{js} | \mathbf{X}_i) \mathbf{x}_{js}\right].\end{aligned}$$

- The crux is to ask which pairs are uncorrelated: $E(u_{it} u_{ij} | \mathbf{X}) = 0$?
- **FE.3:** $E(\ddot{u}_{it} \ddot{u}_{js} | \mathbf{X}_i) = \sigma^2$ only if $i = j \wedge s = t$ (and 0 otherwise).
- **Independent individuals:** $E(\ddot{u}_{it} \ddot{u}_{js} | \mathbf{X}_i) = 0$ when $i \neq j$.
- **Serial correlation** within individuals: $E(\ddot{u}_{it} \ddot{u}_{is} | \mathbf{X}_i) \neq 0$ when $t \neq s$.
- **Independently sampled clusters:** Each i belongs to a cluster,

$C_i \in \{1, \dots, M\}$

More thoroughly

- **Suppose** there is a *partition* of the data: $\{\mathcal{G}_g\}_{g=1}^G$.
 - Each observation is in precisely one “cluster”: $(i, t) \in \mathcal{G}_g$, and $\mathcal{G}_g \cap \mathcal{G}_{g'} = \emptyset$.
- **Suppose** clusters are drawn independently, but errors from the same cluster may be correlated.
- **Then rewrite**

$$\sum_i \sum_t \mathbf{x}'_{it} u_{it} = \sum_{g=1}^G \sum_{(i,t) \in \mathcal{G}_g} \mathbf{x}'_{it} u_{it}.$$

- **Our CLT** then works on

$$\frac{1}{\sqrt{G}} \sum_{g=1}^G \mathbf{w}_g \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{B}).$$

- Where $\mathbf{w}_g = \sum_{(i,t) \in \mathcal{G}_g} \mathbf{x}_{it} u_{it}$.
- **The variance, \mathbf{B} ,** is then

$$\begin{aligned} V(\mathbf{w}_g) &= \sum_{(i,t) \in \mathcal{G}_g} \sum_{(j,s) \in \mathcal{G}_g} \text{Cov}(\mathbf{x}'_{it} u_{it}, \mathbf{x}'_{js} u_{js}) \\ &= \sum_{(i,t) \in \mathcal{G}_g} \sum_{(j,s) \in \mathcal{G}_g} E \left[\mathbf{x}'_{it} E(u_{it} u_{js} | \mathbf{X}) \mathbf{x}_{is} \right]. \end{aligned}$$

Cluster-robust std.err. (p. 865)

If $G \rightarrow \infty$ but the # of i in each g is fixed,

$$\widehat{\text{Avar}}(\beta_{POLS}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{g=1}^G \mathbf{X}'_g \hat{\mathbf{u}}_g \hat{\mathbf{u}}'_g \mathbf{X}_g \right) (\mathbf{X}'\mathbf{X})^{-1}$$

- **In practice:** applied empirical papers *always* use clustered std.errs.
- **Tendency** that clustered std.errs. are **bigger**.
 - If there is correlation within clusters \Rightarrow only CRSE are consistent!
 - If there is **none** \Rightarrow the difference between CRSE and typical SE can be **substantial**!
- \Rightarrow With $N \geq 1$ mio., $\text{se}(\hat{\beta})$ can easily increase 10 times by clustering on municipality!
- **What to cluster on?** Not too few clusters – need $G \rightarrow \infty$.
 - Typical: Individual, firm, municipality, market, year-month, etc.

Towards Visual Intuition: Blocks of $E(\mathbf{u}\mathbf{u}'|\mathbf{X})$

- The FE estimator is

$$\hat{\beta}_{FE} = \beta + (\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'\ddot{\mathbf{u}}.$$

- If a CLT applies, then it always holds:

$$\text{Avar}(\hat{\beta}_{FE}) = [E(\ddot{\mathbf{X}}'\ddot{\mathbf{X}})]^{-1} E[\ddot{\mathbf{X}}'E(\ddot{\mathbf{u}}\ddot{\mathbf{u}}'|\ddot{\mathbf{X}})\ddot{\mathbf{X}}] [E(\ddot{\mathbf{X}}'\ddot{\mathbf{X}})]^{-1}.$$

- **Problem:** $E(\ddot{\mathbf{u}}\ddot{\mathbf{u}}'|\ddot{\mathbf{X}})$ is $NT \times NT$: way too many parameters \Rightarrow structure is needed.
 - IID over i and $t \Rightarrow E(\ddot{\mathbf{u}}\ddot{\mathbf{u}}'|\ddot{\mathbf{X}}) = \sigma^2 \mathbf{I}_{NT}$ (fully diagonal)
 - IID over $i \Rightarrow E(\ddot{\mathbf{u}}\ddot{\mathbf{u}}'|\ddot{\mathbf{X}})$ has N identical blocks of size $T \times T$
 - IID over G clusters $\Rightarrow E(\ddot{\mathbf{u}}\ddot{\mathbf{u}}'|\ddot{\mathbf{X}})$ has G blocks

Figure 1: $\text{Cov}(u_{it}, u_{js}) = 0$ if $i \neq j$ or $s \neq t$

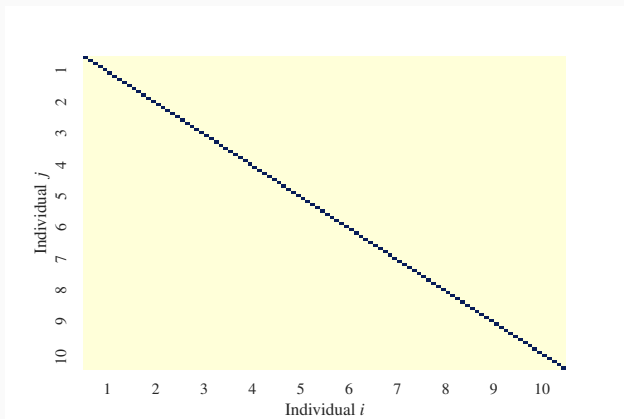
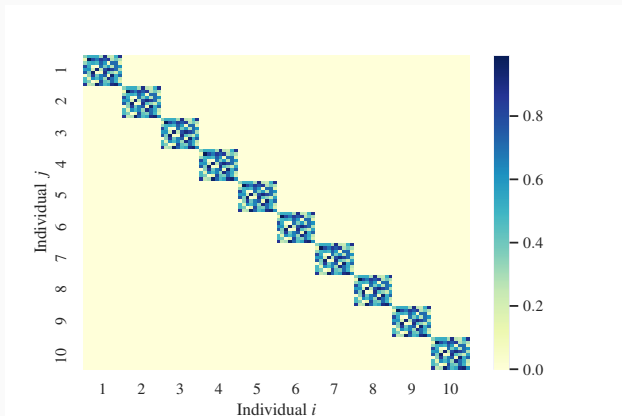


Figure 2: $\text{Cov}(u_{it}, u_{js})$



- The N blocks are complex $T \times T$ matrices, but they are identical

Figure 3: $\text{Cov}(u_{it} + c_i, u_{js} + c_j)$

