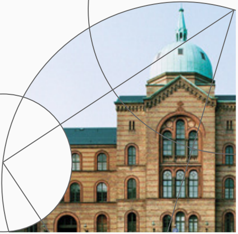
Linear Models for Panel Data

Advanced Microeconometrics

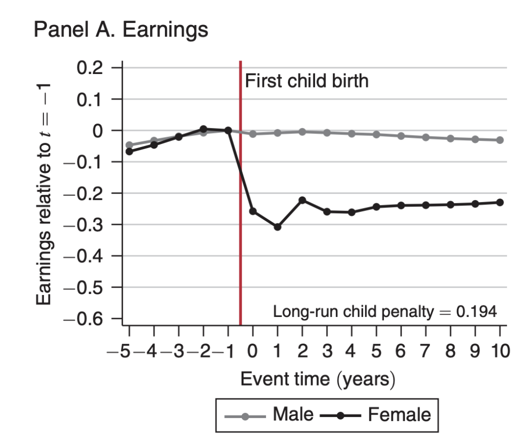
Anders Munk-Nielsen

2022



Introduction

Child Penalty



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

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

Blitz Discussion

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
we are interested in variation of wages within ind that have children not those do not have children
2. Based on the graph, what are β_1 and β_2 approximatey? 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 *inutitive* (meaning: real-world) explanations for correlation (positive or negative) between c_i and
   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

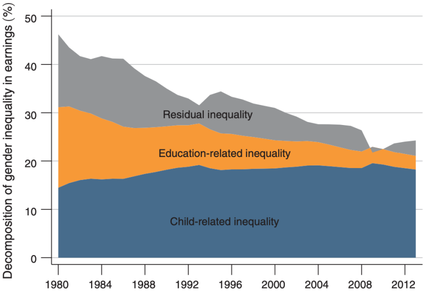


FIGURE 5. DECOMPOSING GENDER INEQUALITY IN EARNINGS

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	$\div$	$\div$	low	$\div$	✓	$\mathbb{R}$	✓
II	LASSO	$\div$	✓	high	✓	$\div$	$\mathbb{R}$	$\div$
IV	Probit	✓	✓	low	✓	✓	$\{0,1\}$	$\div$
	Tobit	✓	✓	low	✓	✓	$[0;\infty)$	$\div$
	Logit	✓	✓	low	✓	✓	$\{1,2,\ldots,J\}$	$\div$
	Sample selection	✓	✓	low	✓	✓	$\mathbb{R}$ and $\{0,1\}$	$\div$
	Simulated Likelihood	✓	✓	low	✓	✓	Any	✓
	Quantile Regression	$\div$	✓	(low)	✓	✓	$\mathbb{R}$	$\div$
	Non-parametric	✓	(✓)	∞	$\div$	$\div$	$\mathbb{R}$	$\div$

Outline

Panel model

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

			Estimator				
Assumptions	Rank	Description	POLS	BE	RE	FE	FD
$\mathrm{E}(u_{it}\vert\mathbf{x}_{it}, c_i) = 0,\ c_i = c,$	$\mathbf{X}'\mathbf{X}$	No individual effects	$\checkmark + \star$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
$\mathrm{E}(u_{it}\vert\mathbf{X}_i, c_i) = 0,\ c_i \sim \mathrm{IID}(0, \sigma_c^2),$	$\mathbf{X}'\mathbf{X}$	Random effects	$\checkmark$	$\checkmark$	$\checkmark + \star$	$\checkmark$	$\checkmark$
$\mathrm{E}(u_{it}\vert\mathbf{X}_i, c_i) = 0,\ \mathrm{E}(c_i\mathbf{x}_{it}) \neq 0,$	$\ddot{\mathbf{X}}'\ddot{\mathbf{X}}$	Fixed effects	$\div$	$\div$	$\div$	$\checkmark$	$\checkmark$
$\mathrm{E}(u_{it}\vert\mathbf{X}_i, c_i) = 0,\ \mathrm{E}(c_i\mathbf{x}_{it}) \neq 0,$ and u_{it} IID	$\ddot{\mathbf{X}}'\ddot{\mathbf{X}}$	Fixed effects	$\div$	$\div$	$\div$	$\checkmark + \star$	$\checkmark$
$\mathrm{E}(u_{it}\vert\mathbf{X}_i, c_i) = 0,\ \mathrm{E}(c_i\mathbf{x}_{it}) \neq 0$ and Δu_{it} IID	$\Delta\mathbf{X}'\Delta\mathbf{X}$	Fixed effects	$\div$	$\div$	$\div$	$\checkmark$	$\checkmark + \star$
$\mathrm{E}(u_{it}\vert\mathbf{x}_{it}, c_i) \neq 0$	–	Endogeneity	$\div$	$\div$	$\div$	$\div$	$\div$

($\div$ = inconsistent, $\checkmark$ = consistent, $\star$ = efficient)

Notation

- **Average:** a "bar" denotes an average,

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

- **Transformations:**
  - Demeaning:

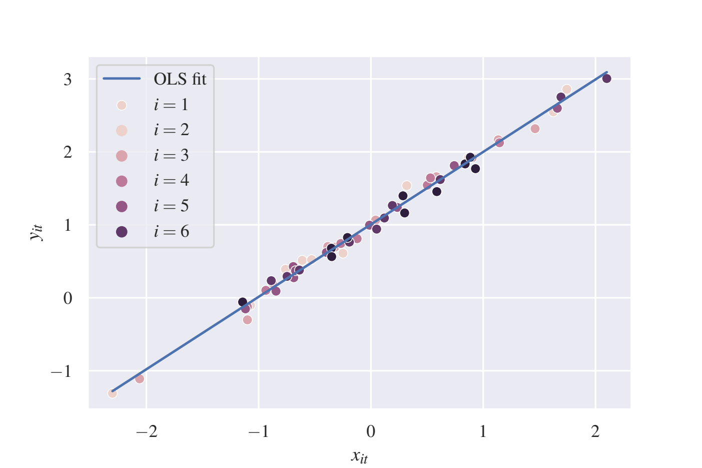
$$\ddot{\mathbf{x}}_{it} \equiv \mathbf{x}_{it} - \bar{\mathbf{x}}_i,$$

  - First-differences:

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

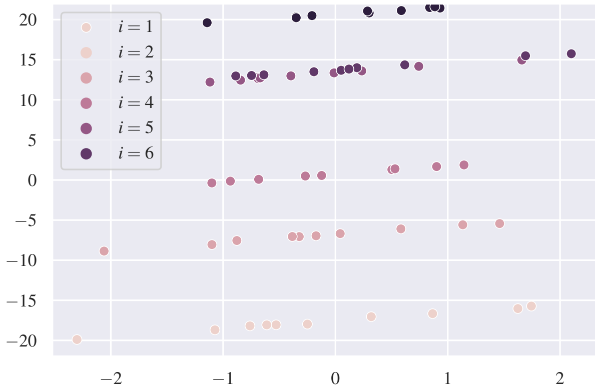
  - ***ALWAYS*** define explicitly what you mean by mathematical symbols.
- **Estimator:** A "hat" denotes an estimator, e.g. $\hat{\boldsymbol{\beta}}$ is an estimate of $\boldsymbol{\beta}$.

No Unobserved Effects: $c_i = 0$

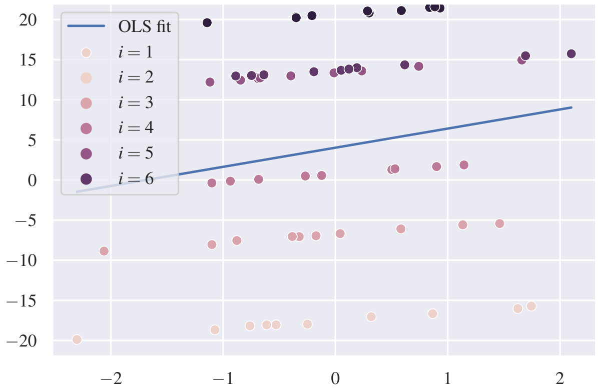


Zooming out on y

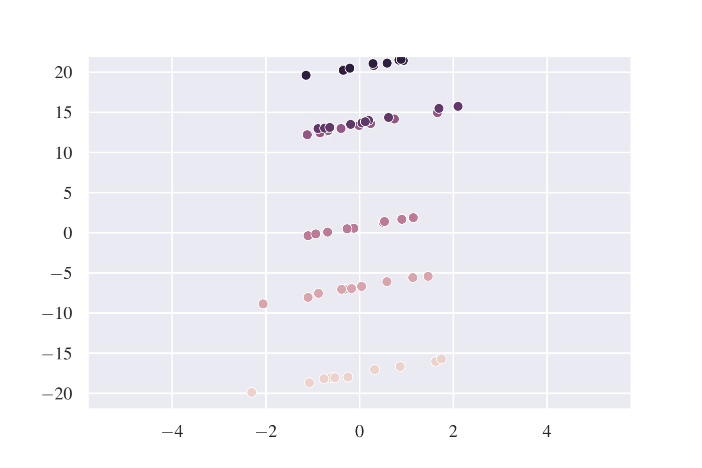
Adding an IID c_i: $\mathrm{E}(\mathbf{x}_i' c_i) = 0$ (Random Effects)



Adding an IID c_i: $E(x_i' c_i) = 0$ – pooled OLS

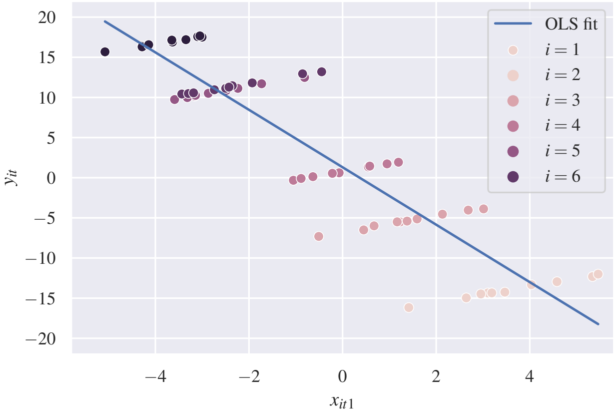


Zooming out on x



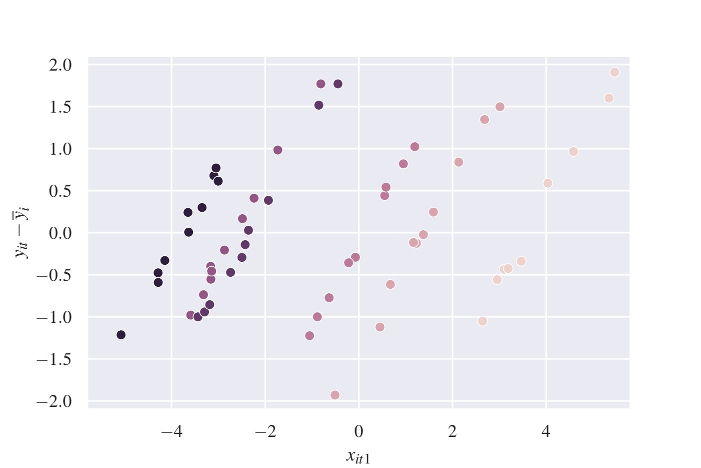
Allowing Endogeneity: $E(x_i' c_i) \neq 0$ (Fixed Effects)

Pooled OLS is inconsistent

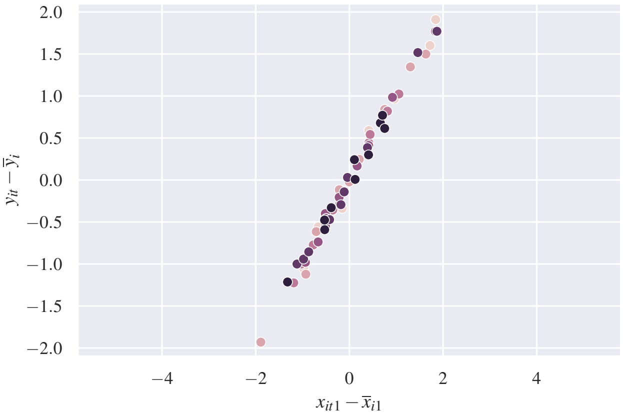


Demeaning y_{it}

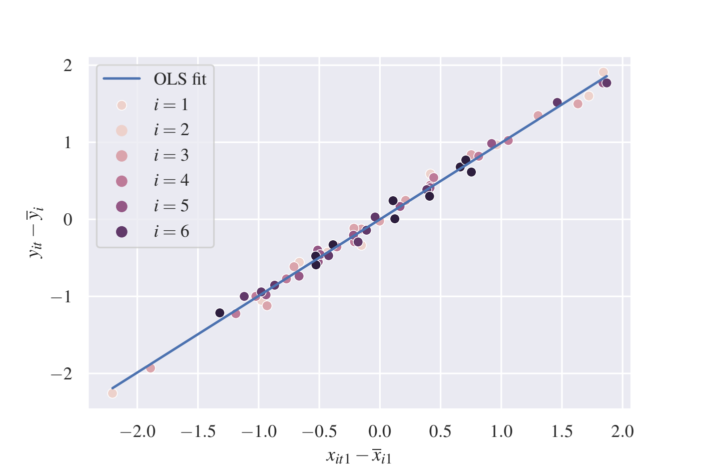
Demeaning y_{it}... and zooming



Also demeaning x_{it}



POLS on Demeaned Data



Model

General linear panel data model

General Model

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

- **Suppose** we assume $u_{it} \sim \text{IID}(0, \sigma_u^2)$
  (i.e. u_{it} independent of all variables, so in particular $\text{E}(u_{it}|\mathbf{x}_{it}, c_{it}, \boldsymbol{\beta}_{it}) = 0$)
  - **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
  - $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}$ (common slope), rules out many interesting effects
  - $c_{it} = c_i$ (time-constant individual specific effect)

Linear panel data model

Model

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

- **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 $\mathrm{E}(c_i) = c$ and have no constant in $\mathbf{x}_{it}$

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

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

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

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

Assumptions

Assumption: strict (or strong) exogeneity (SE)

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

Assumption: contemporaneous exogeneity (CE)

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

- 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: $\mathrm{E}(u_{it}|\mathbf{x}_{it}) = \mathbf{0}$ stricter
    - Unconditional: $\mathrm{E}(\mathbf{x}_{it}u_{it}) = \mathbf{0}$ 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** $\mathrm{E}(y_{it}|\mathbf{x}_{it})$ for different values of $\mathbf{x}_{it}$
  - That is, we can maybe hope to estimate

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

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

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

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

Proving Inconsistency

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

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

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

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

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

Plan

- **Now:** different strategies based on $\mathrm{E}(\mathbf{x}c_i)$.
- Under **exogeneity** of the individual effect, $\mathrm{E}(\mathbf{x}c_i) = \mathbf{0}$:
  - Pooled OLS is consistent
  - Random Effects Estimator is efficient
- Under **endogeneity**, $\mathrm{E}(\mathbf{x}c_i) \neq \mathbf{0}$,
  - 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

Linear panel data model

Model

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

- **Assume:**
  - Strict exogeneity: $\mathrm{E}(u_{it}|\mathbf{X}_i, c_i) = 0$ $(\mathbf{X}_i \equiv (\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}))$
  - Rank condition: (to be specified)
- **Idea:** perform *within transformation*,

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

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

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

- **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 Estimator

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 $\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{\mathbf{u}}_{it}$ ...
  - ... check conditions for OLS on this:
- **Exogeneity:** $\mathrm{E}(\ddot{u}_{it}\ddot{\mathbf{x}}_{it}) = \mathbf{0}$.
- **Rank condition:** $\ddot{\mathbf{X}}'\ddot{\mathbf{X}}$ must be invertible:

FE Assumptions

FE Assumptions for Consistency

1. $\mathrm{E}(\ddot{\mathbf{x}}_{it}\ddot{u}_{it}) = \mathbf{0}$
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 ($\mathrm{E}(u_{it}|\mathbf{x}_{it}) = 0$) is insufficient

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

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

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

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

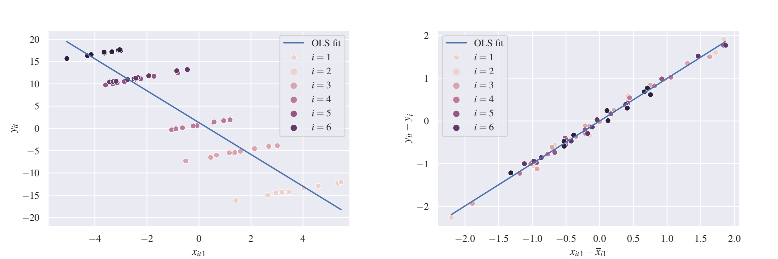
FE Assumptions

FE Assumptions for Consistency

1. $\mathrm{E}(\ddot{\mathbf{x}}_{it}\ddot{u}_{it}) = \mathbf{0}$
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 $\mathrm{Corr}(\mathbf{x}_{it}, c_i)$
- **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

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

- **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 $N \to \infty$ but $T \not\to \infty$, $\dim(\beta)$ 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:**

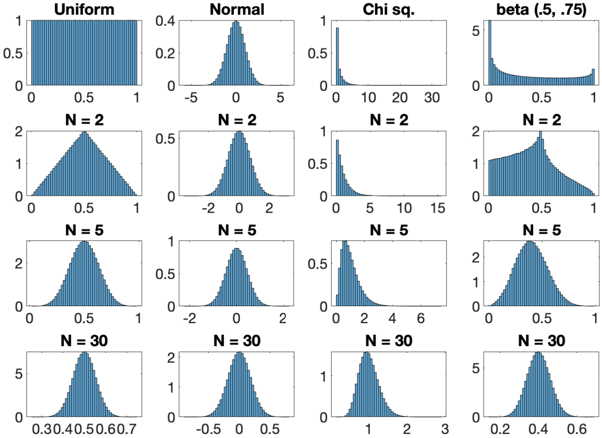
$$\begin{aligned}\hat{\beta}_{FE} &= (\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'(\ddot{\mathbf{X}}\beta + \ddot{\mathbf{u}}) \\ &= \beta + (\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'\ddot{\mathbf{u}}\end{aligned}$$

- **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: $\mathbf{A} \equiv \mathrm{E}(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i)$ ($\ddot{\mathbf{X}}_i$ 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 $\mathrm{E}(\mathbf{w}_i) = \mathbf{0}_{1\times K}$, then

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

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

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

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

Variance

- **We showed:**

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

- **Law of Iterated Expectations (LIE)** giver

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

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

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

Assumption FE.3 (Homoskedasticity)

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

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

Variance

Asymptotically, under FE.1 and FE.2:

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

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

Panel-robust Variance Estimator

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

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

Variance Estimator under Homoskedasticity (FE.3)

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

Comment on $\hat{\sigma}_u^2$

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

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

- **Usual df. correction:** "$-K$": for efficiency, no effect asymptotically (as $N \to \infty$)
  - 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**

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

  - **The crux is** to ask which pairs are uncorrelated: $\mathrm{E}(u_{it}u_{ij}|\mathbf{X}) = 0$?
- **FE.3:** $\mathrm{E}\big(\ddot{u}_{it}\ddot{u}_{js}\big|\mathbf{X}_i\big) = \sigma^2$ only if $i = j \wedge s = t$ (and 0 otherwise).
- **Independent individuals:** $\mathrm{E}\big(\ddot{u}_{it}\ddot{u}_{js}\big|\mathbf{X}_i\big) = 0$ when $i \neq j$.
- **Serial correlation** within individuals: $\mathrm{E}\big(\ddot{u}_{it}\ddot{u}_{is}\big|\mathbf{X}_i\big) \neq 0$ when $t \neq s$.
- **Independently sampled *clusters***: Each i belongs to a cluster, $\mathcal{C}_g \subseteq \{1, \dots, N\}$

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**, **B**, is then

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

Cluster-robust std.err.

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

If $G \to \infty$ 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., se($\hat{\beta}$) can easily increase 10 times by clustering on municipality!
- **What to cluster on?** Not too few clusters – need $G \to \infty$.
  - Typical: Individual, firm, municipality, market, year-month, etc.

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

- **The FE estimator** is

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

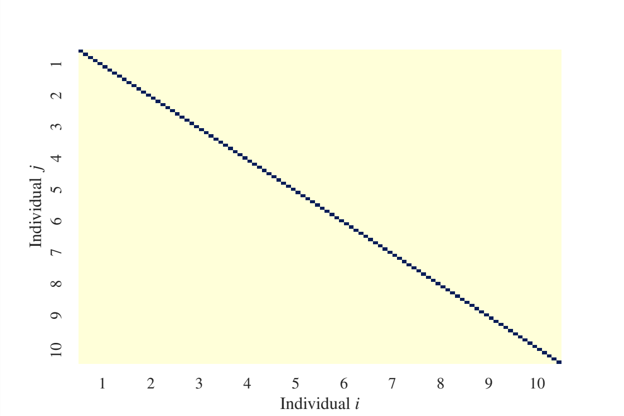
- **If** a CLT applies, then it always holds:

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

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

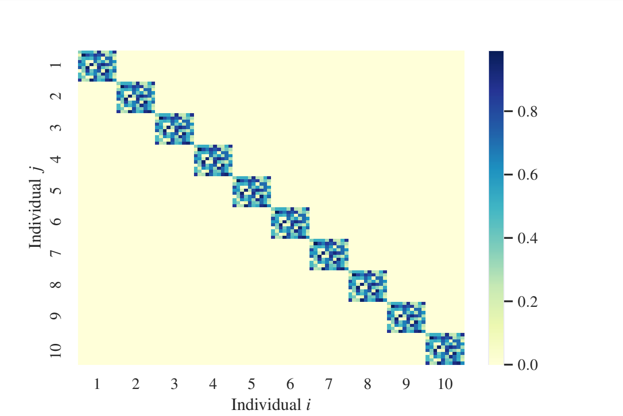
$\mathrm{E}(\ddot{u}\ddot{u}')$ **with** FE.3

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



$\mathrm{E}(\ddot{u}\ddot{u}')$ without FE.3

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



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

RE model: Next lecture

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