



# Linear Models for Panel Data

## Advanced Microeconometrics

---

Anders Munk-Nielsen

2022



# Introduction

---

# Example: Wages and schooling

## Example: Wages and schooling

$$w_{it} = \beta_0 + \beta_1 e_{it} + c_i + u_{it},$$

where  $w_{it}$  is wage and  $e_{it}$  is years of schooling.

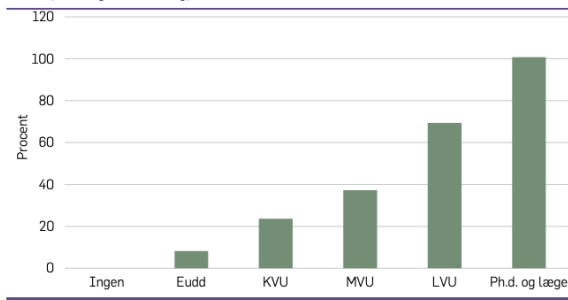
- we can **compute** the average wage among, say, people with  $e_{it} = 10$  years of schooling,  $\hat{E}(w_{it}|e_{it} = 10)$
- **Intuitively**, you can think of the POLS estimator as finding  $\hat{\beta}_1$  by seeing how  $\hat{E}(w_{it}|e_{it})$  varies with *educ*.
- **The problem:** what if high-IQ people (high  $c_i$ , and high *wage*; regardless of *educ*<sub>*i*</sub>) take more education (*educ*)
  - Mathematically:  $\text{Cov}(e_{it}, c_i) > 0$
- **Decomposing** how  $\hat{E}(w_{it}|e_{it})$  increases with  $e_{it}$ :
  - one part is the value of education:  $\beta_1 e_{it}$
  - one part is the *type composition* changing: higher IQ among higher educated.

# Example: Wages and schooling

Estimates from regression of log wage on dummies for education types (and sector dummies and experience)

FIGUR 2.1

Lønafkast pr. uddannelseskategori på hele arbejdsmarkedet set i forhold til lønnen for ansatte uden uddannelse, 2011. Opgjort for standardberegnet timefortjeneste (ekskl. genebetaling). Procent.



## Education: Signaling or Skill Building

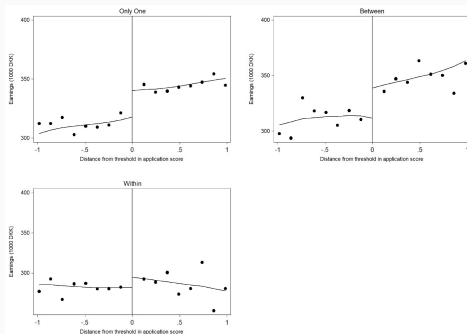
Clearly, highly educated earn more. Pick a side randomly and argue which is more important for economists:

1. **Skills:** education teaches analytical and abstract thinking which heightens productivity. (E.g. Scandinavian school reforms)
2. **Signaling:** smart people take a tough education to show firms how productive they will be. (E.g. almost nobody remembers trigonometry)

Relate the discussion to the linear model below.

Check out the discussion here: [youtu.be/MvWnyUT7vPk](https://youtu.be/MvWnyUT7vPk)

# Admission Cutoffs



Cool empirical strategy: strict cutoff in admission for GPA

- $\Rightarrow$  *Regression Discontinuity* design.

Source: *Daly, Jensen & le Maire (2022; Labour Economics)*.

# 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	✓	(✓)	$\infty$	÷	÷	$\mathbb{R}$	÷

## Panel model

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

Assumptions	Rank	Model name	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	÷	÷	÷	✓+★	✓
$E(u_{it} \mathbf{X}_i, c_i) = 0, E(c_i\mathbf{x}_i) \neq 0 \text{ and } u_{it} - u_{it-1} \text{ IID}$	$\Delta\mathbf{X}'\Delta\mathbf{X}$	Fixed effects	÷	÷	÷	✓	✓+★
$E(u_{it} \mathbf{x}_{it}, c_i) \neq 0$	–	Endogeneity	÷	÷	÷	÷	÷

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



# First Differences

---

# First Differences

- **Now**, consider the first-differences transformed model

$$\begin{aligned}y_{it} - y_{it-1} &= (\mathbf{x}_{it} - \mathbf{x}_{it-1})\beta + \overbrace{c_i - c_i}^{=0} + u_{it} - u_{it-1} \\ \Delta y_{it} &= \Delta \mathbf{x}_{it}\beta + \Delta u_{it}\end{aligned}$$

- **Voila!** Big beautiful success.
- **Implementation:** transform data and run POLS of  $\Delta \mathbf{Y}$  on  $\Delta \mathbf{X}$
- **Assumptions:** Translate between assumptions on raw vs. transformed dataset.
  - Analogously to what we did for FE.

# First Differences Estimator

## First Differences (FD) Estimator

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

## FD Assumptions for Consistency

1.  $E(\Delta \mathbf{x}_{it} \Delta u_{it}) = \mathbf{0}$ ,
2.  $E(\Delta \mathbf{X}' \Delta \mathbf{X})$  must have full rank

(Simply the POLS assumptions written for the transformed data.)

## FD Assumptions for Consistency

1.  $E(\Delta \mathbf{x}_{it} \Delta u_{it}) = \mathbf{0}$ ,
2.  $E(\Delta \mathbf{X}' \Delta \mathbf{X})$  must have full rank

Translating assumptions on transformed to raw variables

- **Exogeneity:**

$$E(\Delta \mathbf{x}_{it} \Delta u_{it}) = E(\mathbf{x}_{it} u_{it}) - E(\mathbf{x}_{it-1} u_{it}) + E(\mathbf{x}_{it} u_{it-1}) + E(\mathbf{x}_{it-1} u_{it-1})$$

- Again, **contemporaneous** exog. is **insufficient**
- **Strict** exogeneity is, since  $E(u_{it} | \mathbf{x}_i) = 0 \Rightarrow E(\mathbf{x}_{it-1} u_{it}) = 0$ .
- **Sufficient:**  $E(u_{it} | \mathbf{x}_{it+1}, \mathbf{x}_{it}, \mathbf{x}_{it-1}) = 0$ .
- **Rank condition:** Requires time-series variation.
  - Same as for FE.

- **Similar** argumentation to the FE case yields:

## Panel-robust variance estimator

$$\widehat{\text{Avar}}(\hat{\beta}_{FD}) = (\sum_i \Delta \mathbf{X}_i' \Delta \mathbf{X}_i)^{-1} \left( \sum_i \Delta \mathbf{X}_i' \widehat{\Delta \mathbf{u}}_i \widehat{\Delta \mathbf{u}}_i' \Delta \mathbf{X}_i \right) (\sum_i \Delta \mathbf{X}_i' \Delta \mathbf{X}_i)^{-1}$$

where  $\widehat{\Delta \mathbf{u}}_i \equiv \Delta \mathbf{y}_i - \Delta \mathbf{X}_i \hat{\beta}_{FD}$  ( $\Delta \mathbf{X}_i$  is  $T - 1 \times K$ )

## Variance estimator for IID errors

Assuming that  $\Delta u_{it}$  are IID (implying that  $u_{it}$  is a random walk!),

$$\widehat{\text{Avar}}(\hat{\beta}_{FD}) = \hat{\sigma}_u^2 (\sum_i \Delta \mathbf{X}_i' \Delta \mathbf{X}_i)^{-1},$$

where

$$\hat{\sigma}_u^2 = \frac{1}{NT - N - K} \sum_i \sum_t \widehat{\Delta u}_{it}^2.$$

How do we choose between FE and FD?

- **Identical** when  $T = 2$ , both numerically and algebraically:
  - E.g.:  $\ddot{y}_{i2} = y_{i2} - \frac{y_{i1} + y_{i2}}{2} = \frac{1}{2}\Delta y_{i2}$ , and  $\ddot{y}_{i1} = -\frac{1}{2}\Delta y_{i1}$ .
- **Consistency:**
  - FE: requires *strict* exogeneity,  $E(u_{it} | \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}) = 0$
  - FD: only requires  $E(u_{it} | \mathbf{x}_{it+1}, \mathbf{x}_{it}, \mathbf{x}_{it-1}) = 0$ , which is *implied* by *strict exogeneity*.
- **Efficiency:** in POLS  $\Rightarrow$  occurs when all errors are IID.
  - FD: efficient if  $\Delta u_{it}$  is IID i.e.  $u_{it}$  is a unit root.
  - FE: efficient if  $u_{it}$  is IID i.e.  $u_{it}$  serially uncorrelated,
    - Why  $u_{it}$  and not  $\ddot{u}_{it}$ ? ... next slide...

- **Question:** Why does FE.3 assume  $u_{it}$  is IID, and not  $\ddot{u}_{it}$ ?
- **First:** note that  $\ddot{u}_{it}$  can never be IID
  - $\ddot{u}_{it} \equiv u_{it} - \bar{u}_i$ : so  $\ddot{u}_{it}$  and  $\ddot{u}_{is}$  share the component  $\bar{u}_i$ .
- **Second:** Recall the “demeaning” matrix,  $\mathbf{Q}_T$ , s.t.  $\ddot{\mathbf{u}}_i = \mathbf{Q}_T \mathbf{u}_i$ .
  - Note:  $\mathbf{Q}_T \mathbf{Q}_T = \mathbf{Q}_T$  (idempotent, symmetric)
  - Next:  $\ddot{\mathbf{X}}_i' \ddot{\mathbf{u}}_i = (\mathbf{Q}_T \mathbf{X}_i)' \mathbf{Q}_T \mathbf{u}_i = \mathbf{X}_i' \mathbf{Q}_T \mathbf{u}_i = \ddot{\mathbf{X}}_i' \mathbf{u}_i$
- **FE** takes the asymptotic form

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

- **Conclusion:** Unweighted estimation on demeaned data (i.e. FE) is efficient when  $E(\mathbf{u}_i \mathbf{u}_i' | \ddot{\mathbf{X}}_i)$  is diagonal, i.e. when  $u_{it}$  are IID.

# Random Effects Model

---



# Between Estimator

- **Stepping stone:** the Between Estimator (BE).
- **New transformation:** take the time-average,

$$\bar{y}_i = \bar{\mathbf{x}}_i \beta + c_i + \bar{u}_i,$$

where  $\bar{\mathbf{x}}_i \equiv T^{-1} \sum_t \mathbf{x}_{it}$  is  $K \times 1$ .

- **Failure:** does *not* remove  $c_i$ .
- **Between (BE) Estimator:** Regress  $\bar{\mathbf{Y}}$  on  $\bar{\mathbf{X}}$ .
- **Does it work?** Check POLS assumptions on transformed data
- **Exogeneity:** Requires  $E[\bar{\mathbf{x}}'_i(c_i + \bar{u}_i)] = \mathbf{0}_{K \times 1}$ .

## BE Assumptions

1.  $E(\bar{\mathbf{x}}_i \bar{u}_i) = \mathbf{0}$  for all  $t, s$ , and  $E(\bar{\mathbf{x}}_i c_i) = \mathbf{0}$ ,
2.  $\bar{\mathbf{X}}' \bar{\mathbf{X}}$  must have full rank.

## Translating to model primitives

- **Exogeneity** sufficient conditions are
  - Strict exogeneity:  $E(u_{it} | \mathbf{x}_i) = 0$  for all  $t$
  - Uncorrelated individual effects:  $E(\mathbf{x}_{it} c_i) = \mathbf{0}$  for all  $t$ 
    - $\Rightarrow$  i.e. we are in the random effects model.
- **Rank condition:** unlike FE or FD, we can now allow regressors with zero time-variation.
  - **Bonus question:** can you think of a variable allowed by BE but not FE rank condition?  
gender, race etc  
anything constant over time as subtracting the mean of regressors would always result in columns of zero  $\rightarrow$  breaking the rank condition

# Random Effects Model

## Model

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}, \quad c_i \sim \text{IID}(0, \sigma_c^2), u_{it} \sim \text{IID}(0, \sigma_u^2).$$

## Assumptions

- RE 1(a):  $E(u_{it}|\mathbf{X}_i) = 0$
- RE 1(b):  $E(c_i|\mathbf{x}_{it}) = 0$
- RE 2:  $E(\mathbf{X}_i'\Omega^{-1}\mathbf{X}_i)$  has full rank, where  $\Omega \equiv E(\mathbf{v}_i\mathbf{v}_i')$
- RE 3(a):  $E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{X}_i, c_i) = \sigma_u^2\mathbf{I}_T$
- RE 3(b):  $E(c_i^2|\mathbf{x}_i) = \sigma_c^2$

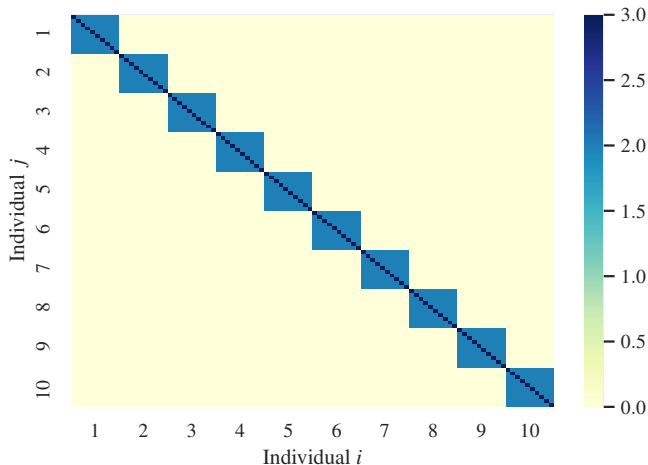
# Random Effects Assumptions

- **POLS consistent:** with both  $E(\mathbf{x}_{it}u_{it}) = \mathbf{0}$  and  $E(\mathbf{x}_{it}c_i) = \mathbf{0}$ , we have consistency of pooled OLS.
  - RE furthermore requires *strict* (and not just contemporaneous) exogeneity.
- **Claim:** POLS is not efficient when  $\sigma_c > 0$ .
- **Proof:** POLS is efficient under IID errors. In particular,  $\text{Cov}(v_{it}, v_{js} | \mathbf{X}) = 0$  for all  $i, j, t, s$ .
- Consider the composite error term,  $v_{it} \equiv c_i + u_{it}$
- The covariance is

$$\text{Cov}(v_{it}, v_{js} | \mathbf{X}) = \begin{cases} 0 & \text{if } i \neq j \text{ by independence over } i, \\ \sigma_c^2 & \text{if } i = j \text{ and } t \neq s, \\ \sigma_c^2 + \sigma_u^2 & \text{if } i = j \text{ and } t = s. \end{cases}$$

- I.e.  $V(\mathbf{v} | \mathbf{X})$  is *block diagonal* with  $N$  consecutive  $T \times T$  blocks.
- **QED.**

# Visualizing the composite error covariance: $\text{Cov}(c_i + u_{it}, c_j + u_{js})$



# Random Effects Assumptions

- **Rank condition:** assumes that  $\mathbf{X}'\mathbf{X}$  is invertible.
  - Recall: FE/FD assumes that the *transformed* matrices are invertible.
  - **Implication:** time-invariant variables are *allowed*.
    - **Implicit assumption:** they are independent of  $c_i$ .
- **Efficiency:** our IID assumptions imply **homoskedasticity** for both  $u$  and  $c$ :
  - $E(\mathbf{u}_i \mathbf{u}_i' | \mathbf{x}_i, c_i) = \sigma_u^2 \mathbf{I}_{T \times T}$ : i.e. homoskedasticity and serial uncorrelatedness of  $u_{it}$ 
    - where  $\mathbf{I}_{T \times T}$  is the identity matrix.
  - $E(c_i^2 | \mathbf{x}_i) = \sigma_c^2$
- **Then** the variance of  $v_{it} \equiv c_i + u_{it}$  becomes

$$\Omega \equiv E(\mathbf{v}_i \mathbf{v}_i' | \mathbf{x}_i) = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \cdots & \sigma_c^2 \\ \vdots & \ddots & \vdots \\ \sigma_c^2 & \cdots & \sigma_c^2 + \sigma_u^2 \end{pmatrix} = \sigma_c^2 \mathbf{1}_T \mathbf{1}_T' + \sigma_u^2 \mathbf{I}_{T \times T}$$

- **As shown:** error terms are correlated across rows when  $\sigma_c > 0$ , so POLS is inefficient.
- **However:** we derived the error variance, so we have a recipe for a weighted least squares estimator (which recovers efficiency).
- **Efficient estimation** is performed with weighted least squares (GLS),

$$\hat{\beta}_{GLS} = \left( \sum_{i=1}^N \mathbf{x}_i' \Omega^{-1} \mathbf{x}_i \right)^{-1} \left( \sum_{i=1}^N \mathbf{x}_i' \Omega^{-1} \mathbf{y}_i \right).$$

# Random Effects Estimator

## Random Effects (RE) Estimator

1. **Estimate**  $\sigma_u, \sigma_c$ : Compute the FE and BE estimators, and obtain the residuals

$$\hat{\sigma}_u^2 = \frac{1}{NT - N - K} \sum_i \sum_t (\hat{u}_{it}^{FE})^2 \quad \hat{u}_{it}^{FE} \equiv (y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \hat{\beta}_{FE}$$

$$\hat{\sigma}_c^2 = \frac{1}{N - K - 1} \sum_i (\hat{u}_{it}^{BE})^2 - T^{-1} \hat{\sigma}_u^2, \quad \hat{u}_{it}^{BE} \equiv \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\beta}_{BE}.$$

(note that in the BE,  $\bar{\mathbf{x}}_i$  should contain an intercept (thus  $K + 1$  columns), while  $\mathbf{x}_{it}$  in the FE should not)

2. **Estimate**  $\hat{\beta}_{RE}$  using (F)GLS

$$\hat{\beta}_{RE} = \left( \sum_i \mathbf{x}_i' \hat{\Omega}^{-1} \mathbf{x}_i \right)^{-1} \sum_i \mathbf{x}_i' \hat{\Omega}^{-1} \mathbf{y}_i,$$

$$\text{where } \hat{\Omega} = \hat{\sigma}_c^2 \mathbf{1}_T \mathbf{1}_T' + \hat{\sigma}_u^2 \mathbf{I}_{T \times T}.$$



## Random Effects (RE) Estimator

$$\hat{\beta}_{RE} = \left( \sum_i \mathbf{x}_i \hat{\Omega}^{-1} \mathbf{x}_i' \right)^{-1} \sum_i \mathbf{x}_i \hat{\Omega}^{-1} \mathbf{y}_i,$$

$$\text{where } \hat{\Omega} = \hat{\sigma}_c^2 \mathbf{1}_T \mathbf{1}_T' + \hat{\sigma}_u^2 \mathbf{I}_{T \times T},$$

- **Extremes** wrt.  $\sigma_c$  vs.  $\sigma_u$ :
  - $\sigma_c = 0$  (no random effects):  $\hat{\beta}_{RE} = \hat{\beta}_{POLS}$  since  $\hat{\Omega}$  becomes diagonal.
  - $\sigma_u = 0$  (only random effects):  $\hat{\beta}_{RE} \rightarrow \hat{\beta}_{FE}$ ... easier to see later.
- **Turns out:**  $\hat{\beta}_{RE}$  can also be obtained by running POLS on a transformed dataset.

# Quasi-demeaning

- Define:

$$\lambda \equiv 1 - \sqrt{\frac{\sigma_u^2}{T\sigma_c^2 + \sigma_u^2}} \quad (\text{note that } \lambda \in [0; 1].)$$

- It can be shown that

$$\Omega^{-\frac{1}{2}} = \frac{1}{\sigma_u} (\mathbf{I}_T - \lambda \mathbf{P}_T), \quad \mathbf{P}_T \equiv \frac{1}{T} \mathbf{j}_T \mathbf{j}_T'.$$

- Define  $\mathbf{C}_T \equiv (\mathbf{I}_T - \lambda \mathbf{P}_T)$  ( $T \times T$ )
- **RE Estimator** arises from estimation on the transformed system

$$\mathbf{C}_T \mathbf{y}_i = \mathbf{C}_T \mathbf{X}_i \beta + \mathbf{C}_T \mathbf{v}_i.$$

- **Crucially:** The new errors,  $\check{\mathbf{v}}_i \equiv \mathbf{C}_T \mathbf{v}_i$ , satisfy homoskedasticity:

$$E(\check{\mathbf{v}}_i \check{\mathbf{v}}_i') = \mathbf{C}_T \Omega \mathbf{C}_T = \sigma_u^2 \mathbf{I}_T.$$

# Estimation of $\sigma_c, \sigma_u$ by POLS

- **Challenge:** Obtaining estimates of  $\sigma_c, \sigma_u$ .
- **Simpler problem:** estimating  $\sigma_c$  and  $\sigma_v$ 
  - $\Rightarrow$  then  $\sigma_u^2 = \sigma_v^2 - \sigma_c^2$  under RE.3
- **Estimating  $\sigma_v^2$ :** Use POLS

$$\hat{\sigma}_v^2 = \frac{1}{NT - K} \sum_{i=1}^N \sum_{t=1}^T (\hat{v}_{it}^{POLS})^2,$$

where  $\hat{v}_{it}^{POLS} \equiv y_{it} - \mathbf{x}_{it}\hat{\beta}_{POLS}$ ,  $\hat{\beta}_{POLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ .

- Consistent? Yes, under RE.1
- **Estimating  $\sigma_c^2$ :** Still using POLS residuals
  - Note,  $\Omega$  has elements  $\sigma_c^2$  everywhere in the lower triangle, which has  $T(T-1)\frac{1}{2}$  elements

$$\hat{\sigma}_c^2 = \frac{1}{NT(T-1)\frac{1}{2} - K} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{t-1} \hat{v}_{it}^{POLS} \hat{v}_{is}^{POLS}.$$

# Alternative: Using FE and BE

- **Alternative idea** for estimating  $(\sigma_u, \sigma_c)$ .
  - **Idea:** Note that both FE and BE are consistent under RE1.
- **FE residuals:** information about  $u_{it}$ ,

$$\hat{\sigma}_u^2 = \frac{1}{NT - N - K} \sum_i \sum_t (\hat{u}_{it}^{FE})^2, \quad \hat{u}_{it}^{FE} \equiv (y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \hat{\beta}_{FE}$$

- **BE residuals:** information about  $c_i$ , ...
  - ... but there is still  $T^{-1} \sum_{t=1}^T u_{it}$  in the BE residual.

$$\hat{\sigma}_c^2 = \frac{1}{N - K - 1} \sum_i (\hat{u}_{it}^{BE})^2 - T^{-1} \hat{\sigma}_u^2, \quad \hat{u}_{it}^{BE} \equiv \bar{y}_i - \bar{\mathbf{x}}_i \hat{\beta}_{BE}.$$

# RE Estimator by Quasi-demeaning

## Random Effects (RE) Estimator

1. Estimate the FE and BE estimators, and compute

$$\hat{\sigma}_u^2 = \frac{1}{NT - N - K} \sum_i \sum_t (\hat{u}_{it}^{FE})^2, \quad \hat{u}_{it}^{FE} \equiv (y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \hat{\beta}_{FE}$$

$$\hat{\sigma}_c^2 = \frac{1}{N - K - 1} \sum_i (\hat{u}_{it}^{BE})^2 - T^{-1} \hat{\sigma}_u^2, \quad \hat{u}_{it}^{BE} \equiv \bar{y}_i - \bar{\mathbf{x}}_i \hat{\beta}_{BE}.$$

2. Transform data by quasi-demeaning:  $\check{y}_{it} \equiv y_{it} - \hat{\lambda} \bar{y}_i$  and  $\check{\mathbf{x}}_{it} \equiv \mathbf{x}_{it} - \hat{\lambda} \bar{\mathbf{x}}_i$ , where

$$\hat{\lambda} = 1 - \sqrt{\frac{\hat{\sigma}_u^2}{T \hat{\sigma}_c^2 + \hat{\sigma}_u^2}}.$$

(Note that  $\hat{\lambda} \in [0; 1]$ .)

3. Estimate by POLS on transformed dataset

# RE Interpretation: Revisited

- **RE:** POLS on quasi-demeaned data,  $\check{y}_{it} \equiv y_{it} - \hat{\lambda}\bar{y}_i$  and  $\check{x}_{it} \equiv x_{it} - \hat{\lambda}\bar{x}_i$ , where

$$\hat{\lambda} = 1 - \sqrt{\frac{\hat{\sigma}_u^2}{T\hat{\sigma}_c^2 + \hat{\sigma}_u^2}} = 1 - \sqrt{\frac{1}{T\frac{\hat{\sigma}_c^2}{\hat{\sigma}_u^2} + 1}}$$

Focus on  $\frac{\sigma_c^2}{\sigma_u^2}$

- **Intuitively:** The RE estimator uses both within and between variation.
- **POLS:**  $\frac{\sigma_c^2}{\sigma_u^2} \rightarrow 0 \Rightarrow \lambda \rightarrow 0$ : no individual effects so demeaning is not necessary.
  - RE and POLS estimates will be close.
- **FE:**  $\frac{\sigma_c^2}{\sigma_u^2} \rightarrow \infty$  or when  $T \rightarrow \infty$ .
  - $\frac{\sigma_c^2}{\sigma_u^2} \rightarrow \infty$ : almost all variation occurs between individuals.
  - $T \rightarrow \infty$ : enough variation within individuals to discard the between variation.

# Variance Estimator

## Variance Estimator

Assuming that  $u_{it}$  and  $c_i$  are IID,

$$\widehat{\text{Avar}}(\hat{\beta}_{RE}) = \hat{\sigma}_u^2 (\sum_i \sum_t \check{\mathbf{x}}'_{it} \check{\mathbf{x}}_{it})^{-1}.$$

## Robust Variance Estimator

$$\widehat{\text{Avar}}(\hat{\beta}_{RE}) = (\sum_i \sum_t \check{\mathbf{x}}'_{it} \check{\mathbf{x}}_{it})^{-1} \sum_i \sum_{t=1}^T \sum_{s=1}^T \check{\mathbf{x}}'_{it} \hat{u}_{it} \hat{u}_{is} \check{\mathbf{x}}_{is} (\sum_i \sum_t \check{\mathbf{x}}'_{it} \check{\mathbf{x}}_{it})^{-1}$$

where  $\hat{u}_{it} \equiv \hat{u}_{it} - \hat{\lambda} \hat{u}_{it}$ , and  $\hat{u}_{it}$  is the RE residual. This allows arbitrary autocorrelation and heteroskedasticity.

- Note: somewhat awkward to use an RE estimator if one fears that RE3 is not satisfied...

## Random Effects Model

Consider the model

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

and assume  $c_i \sim \text{IID}(0, \sigma_c^2)$ ,  $u_{it} \sim \text{IID}(0, \sigma_u^2)$ ,

(meaning that  $\mathbf{x}_{it}$  is independent of both  $c_i$  and  $u_{it}$ ).

## Discuss

Assume that  $T = 1$  and that  $\text{Rank}[E(\mathbf{x}'_{it}\mathbf{x}_{it})] = K$ .

1. Why is  $\hat{\beta}_{POLS}$  consistent? Because  $\mathbf{x}_{it}$  is uncorrelated with  $c_i$  and  $u_{it}$

2. Why can we *not* identify both  $\sigma_c$  and  $\sigma_u$ ? We have no way to separate the individual heterogeneity and "idiosyncratic" errors(?)  
In fact, because we are in a single time period, the idiosyncratic errors are in fact no longer idiosyncratic! The 'noise' do not change over time - we can only observe one time period - so  $c_i + u_{it}$

3. Why would panel data ( $T > 1$ ) help?

This would allow us to identify the variance of the idiosyncratic noise, allowing for noise to vary over time and unit of interests



**FE vs. RE**

---

# Hausman Test

- **Question:** is FE or RE model appropriate?
  - $\Rightarrow$  question posed by the **Hausman Test**.
- **The null hypothesis:** the RE model is appropriate, i.e.

$$\mathcal{H}_0 : \text{RE.1-3 and FE.1-3 hold.}$$

- **Idea:** Under  $\mathcal{H}_0$ :
  - both  $\hat{\beta}_{RE}$  and  $\hat{\beta}_{FE}$  are consistent,
  - both become normal with rate  $\sqrt{N}$ .
- **Result:** because the difference of two normals is normal:

$$\sqrt{N} \left( \hat{\beta}_{RE} - \hat{\beta}_{FE} \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_H).$$

- **Test statistic:** Idea: " $z_k \sim \mathcal{N}(0, \sigma^2) \Rightarrow \sum_{k=1}^K \frac{z_k^2}{\sigma^2} \sim \chi^2(K)$ "

$$H = \left( \hat{\beta}_{RE} - \hat{\beta}_{FE} \right)' \mathbf{V}_H^{-1} \left( \hat{\beta}_{RE} - \hat{\beta}_{FE} \right) \sim \chi^2(\text{dim}(\beta)).$$

# Hausman Test: Variance

- **Missing piece:** How do we compute  $\text{Avar}(\hat{\beta}_{RE} - \hat{\beta}_{FE})$ ?
- **In general,**  $V(X - Y) = V(X) + V(Y) - 2\text{Cov}(X, Y)$
- **Full efficiency case:** it can be shown that when  $\hat{\beta}_{RE}$  is fully efficient,  $\text{Cov}(\hat{\beta}_{RE}, \hat{\beta}_{FE}) = V(\hat{\beta}_{RE})$ , so

$$\text{Avar}(\hat{\beta}_{RE} - \hat{\beta}_{FE}) = \text{Avar}(\hat{\beta}_{FE}) - \text{Avar}(\hat{\beta}_{RE}).$$

- $\mathbf{V}_H$  is always positive definite – in the scalar case,  
 $V(\hat{\beta}_{FE}) > V(\hat{\beta}_{RE})$  both under the null and the alternative.
- However, if  $\hat{\sigma}_u$  estimates come from different models, it can cause trouble.
- **When  $\hat{\beta}_{RE}$  is not fully efficient**, “bootstrapping” can be used.  
*BUT!* Then the variance formula changes...

# Hausman Test: Efficient case

## Hausman Test

The test statistic is

$$H = \left( \hat{\beta}_{RE} - \hat{\beta}_{FE} \right)' \left[ V \left( \hat{\beta}_{FE} \right) - V \left( \hat{\beta}_{RE} \right) \right]^{-1} \left( \hat{\beta}_{RE} - \hat{\beta}_{FE} \right).$$

Under the null hypothesis that assumptions RE.1-3 and FE.1-3 are true, we have

$$H \stackrel{a}{\sim} \chi^2(K),$$

where  $K = \dim(\beta_{FE})$  (number of time-invariant regressors)

# The Bootstrap

---

# The Bootstrap: Motivation

- **Hausman's insight:**  $V(\hat{\beta}_{FE} - \hat{\beta}_{RE})$  takes a particularly simple form under the null.
- **Objection:** What if the null doesn't hold, e.g. due to heteroskedasticity?
  - What if  $V(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \neq V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})$ ?
- **Analytic expressions** may be beyond our capability.
- **More generally, The Bootstrap** procedure provides help in seemingly helpless situations.

# The Bootstrap Procedure

- **Write** our estimator as  $\hat{\beta} = f(\{y_i, \mathbf{x}_i\}_{i=1}^N)$ 
  - It is a function of a dataset

## Bootstrapping

For replications  $r = 1, \dots, R$  do:

- Take  $N$  draws from  $\{1, \dots, N\}$  with replacement, yielding  $\{y_i^{(r)}, \mathbf{x}_i^{(r)}\}_{i=1}^N$ .
  - Note: some individuals,  $i$ , may be drawn multiple times, and some not at all.
- Compute  $\hat{\beta}^{(r)} = f(\{y_i^{(r)}, \mathbf{x}_i^{(r)}\}_{i=1}^N)$

The bootstrapped standard error is then

$$\hat{V}^{Boot}(\hat{\beta}) = \frac{1}{R} \sum_{r=1}^R \left( \hat{\beta}^{(r)} - \bar{\hat{\beta}} \right)^2, \quad \bar{\hat{\beta}} = \frac{1}{R} \sum_{r=1}^R \hat{\beta}^{(r)}.$$

# In Python

```
1 import numpy as np
2 N,K = X.shape
3 for r in range(R):
4     #select rows randomly with replacement
5     ii = np.random.choice(range(N), N, replace=True)
6     X_r = X[ii, :]
7     y_r = y[ii] # assuming y.ndim == 1
8     betas[r, :] = estimate(X_r, y_r)
9
10 # compute empirical covariances and put into a matrix
11 C = np.cov(betas, rowvar=False) # variables are in columns
12 se = np.sqrt(np.diag(C)) # alternative: np.std(betas,0)
```



# Bootstrapping the Hausman Test

- In the Hausman test, we require  $\text{Var}(\hat{\beta}_{RE} - \hat{\beta}_{FE})\dots$
- ... so in each replication, **instead** of computing a single estimate,
  - we we **estimate both**  $\hat{\beta}_{RE}$  and  $\hat{\beta}_{FE}$  and compute the difference
- **From scalar to vector:** Simply compute the empirical Covariance matrix: `np.cov()`
  - Note that `cov` by default expects rows to be variables.