Lecture 2: Linear Model with Panel Data: Fixed Effects and First Differencing

Bertel Schjerning

University of Copenhagen, Department of Economics

January 29, 2026

Plan for Lectures on Linear Models

Lecture 1: Recap, Cross section, Least squares (W.4)

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

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

A Typical Cross Section

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	$x_N^{(2)}$	$x_N^{(3)}$	$\cdots$	$x_N^{(K)}$

Panel Data (Balanced)

i	t	y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$\cdots$	$x^{(K)}$
1	1	y_{11}	$x_{11}^{(1)}$	$x_{11}^{(2)}$	$x_{11}^{(3)}$	$\cdots$	$x_{11}^{(K)}$
1	2	y_{12}	$x_{12}^{(1)}$	$x_{12}^{(2)}$	$x_{12}^{(3)}$	$\cdots$	$x_{12}^{(K)}$
1	3	y_{13}	$x_{13}^{(1)}$	$x_{13}^{(2)}$	$x_{13}^{(3)}$	$\cdots$	$x_{13}^{(K)}$
2	1	y_{21}	$x_{21}^{(1)}$	$x_{21}^{(2)}$	$x_{21}^{(3)}$	$\cdots$	$x_{21}^{(K)}$
2	2	y_{22}	$x_{22}^{(1)}$	$x_{22}^{(2)}$	$x_{22}^{(3)}$	$\cdots$	$x_{22}^{(K)}$
2	3	y_{23}	$x_{23}^{(1)}$	$x_{23}^{(2)}$	$x_{23}^{(3)}$	$\cdots$	$x_{23}^{(K)}$
$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	1	y_{N1}	$x_{N1}^{(1)}$	$x_{N1}^{(2)}$	$x_{N1}^{(3)}$	$\cdots$	$x_{N1}^{(K)}$
N	2	y_{N2}	$x_{N2}^{(1)}$	$x_{N2}^{(2)}$	$x_{N2}^{(3)}$	$\cdots$	$x_{N2}^{(K)}$
N	3	y_{N3}	$x_{N3}^{(1)}$	$x_{N3}^{(2)}$	$x_{N3}^{(3)}$	$\cdots$	$x_{N3}^{(K)}$

Sampling Scheme for (Micro) Panel Data

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

- Observations iid across i.
- Each i observed for T time periods $\{(y_{it}, \mathbf{x}_{it})\}_{t=1}^{T}$.
- May have (within i) dependence across time.

Focus: On asymptotics as $\#i$ grows without bound

- Limits understood as $N \to \infty$ holding T (+other) fixed.
- Implicit assumption for asymptotics to be relevant:

$$N \text{ much larger than } T.$$

- E.g. Danish registers.

Outline

Linear Unobserved Effects Model for Panel Data

Fixed-Effects Methods
- Within Transformation and Estimation
- Identification and Consistency
- Efficiency
- Inference

First-Difference Methods
- First-Difference Transformation and Estimation
- Identification and Consistency
- Efficiency
- Inference

Linear Unobserved Effects Model for Panel Data

Motivation: Why Care About Panel Data?

Object of interest:

To estimate partial effects $\boldsymbol{\beta}$ in population regression function

$$\mathrm{E}\left[y \mid \mathbf{x}, c\right] = \mathbf{x}\boldsymbol{\beta} + c$$

controlling for c.

Problem: c is *unobserved*

Major motivation for using panels:

To account for unobserved individual heterogeneity, c.

Example: Returns to Education

Q: Causal effect of added year of `educ` on `wage`, on average?

We observe `educ` and `wage`.

Regress `wage` on `educ` = success? [D]

E[Wage | Education, . . .]

Linear Unobserved Effects Panel Data Model

Equation of interest in error form

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad t = 1, 2, \ldots, T.$$

If we were to assume

$$\mathrm{E}\left[u_{it} \mid \mathbf{x}_{it}, c_i\right] = 0, \quad t = 1, 2, \ldots, T. \tag{1}$$

we'd get

$$\mathrm{E}\left[y_{it} \mid \mathbf{x}_{it}, c_i\right] = \mathbf{x}_{it}\boldsymbol{\beta} + c_i, \quad t = 1, 2, \ldots, T.$$

- **Note:** Unobserved effect c_i fixed over time.
- (1) = contemporaneous exogeneity conditional on c_i.

Can We Ignore Unobserved Effect?

... and just use (pooled) OLS?

May write model as

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}, \quad t = 1, 2, \ldots, T,$$

where we introduce the composite error

$$v_{it} := c_i + u_{it}.$$

Pooled OLS consistency requires

$$\mathrm{E}\left[\mathbf{x}_{it}' v_{it}\right] = \mathrm{E}\left[\mathbf{x}_{it}' c_i\right] + \mathrm{E}\left[\mathbf{x}_{it}' u_{it}\right] = \mathbf{0} \quad \text{all } t.$$

Relies on $\mathrm{E}\left[\mathbf{x}_{it}' c_i\right] = \mathbf{0}$ all t.

Otherwise, **omitted variable problem**.

Panel Data Estimators and Empirical Strategies

If we suspect $\mathrm{E}\left[\mathbf{x}'_{it} c_i\right] \neq \mathbf{0}$ some t

- Fixed Effects (FE)
- First Differences (FD)

Common feature: Transform data to eliminate c_i.

If we believe $\mathrm{E}\left[\mathbf{x}'_{it} c_i\right] = \mathbf{0}$ all t

- Pooled OLS
- Generalized LS (GLS): Random Effects (for efficiency)

Common feature: Model c_i (cond'l on $\mathbf{x}_{it}$s).

Fixed-Effects Methods

Within Transformation and Estimation

Within Transformation

Recall unobserved effects model

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad t = 1, 2, \ldots, T.$$

Let "bars" denote time averages for given i, e.g.

$$\bar{y}_i := \frac{1}{T}\sum_{t=1}^{T} y_{it} \quad \text{and} \quad \bar{\mathbf{x}}_i := \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}.$$

By time invariance, $\bar{c}_i = c_i$.

Hence

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

Within Transformation

Subtract time-averaged outcome from original:

$$y_{it} - \overline{y}_i = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\boldsymbol{\beta} + (c_i - c_i) + (u_{it} - \overline{u}_i).$$

Eureka! Individual effect cancels out

Let "dots" denote time-demeaning/within transforming, e.g.

$$\ddot{y}_{it} := y_{it} - \overline{y}_i \quad \text{and} \quad \ddot{\mathbf{x}}_{it} := \mathbf{x}_{it} - \overline{\mathbf{x}}_i.$$

Arrive at

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}, \quad t = 1, 2, \ldots, T.$$

Familiar?

Within Transformed Data and OLS

Fixed Effects (FE) = pooled OLS having gone within:

$$\widehat{\beta}_{FE} := \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{y}_{it} \right).$$

Stacking $\ddot{\mathbf{x}}_{it}$s and $\ddot{y}_{it}$s over first t and then i,

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

where

- $\ddot{\mathbf{X}}_i$ is $T \times K$ (presuming all K regressors time-varying),
- $\ddot{\mathbf{y}}_i$ is $T \times 1$,

... and matrices stack these over i.

Discussion

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}, \quad t = 1, 2, \ldots, T.$$

Except for time invariance, have placed no restrictions on c_i.

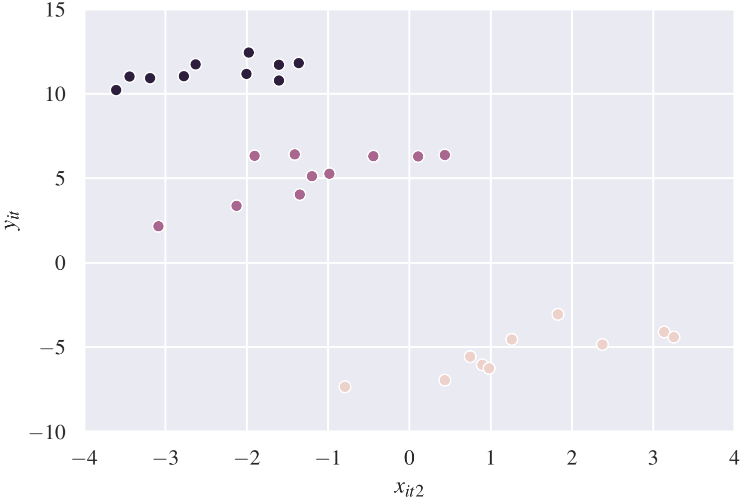
Arbitrary dependence between $\mathbf{x}_{it}$s and c_i allowed.

Remember: Interpretation of $\boldsymbol{\beta}$ comes from original eqn.

How to go within? Multiply each $\mathbf{X}_i$ and $\mathbf{y}_i$ with

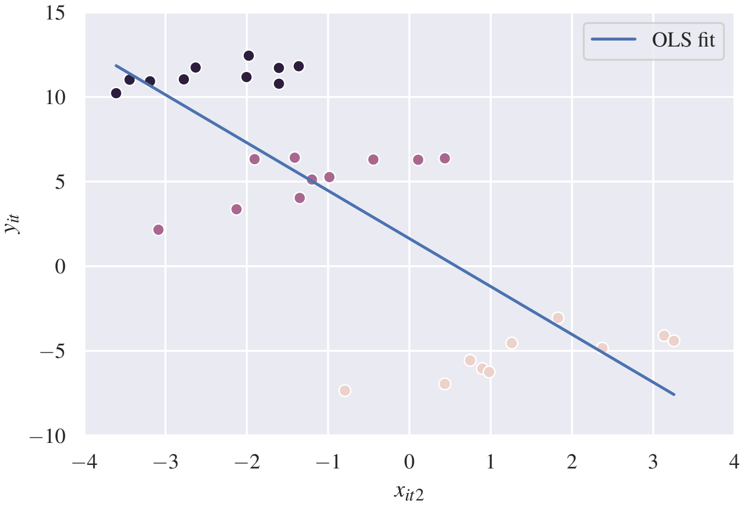
$$\text{Time-demeaner:} \quad \mathbf{Q}_T := \mathbf{I}_T - \frac{1}{T}\mathbf{j}_T\mathbf{j}_T'. \qquad (T \times T)$$

Simulated Data ($N = 3$, $T = 10$)



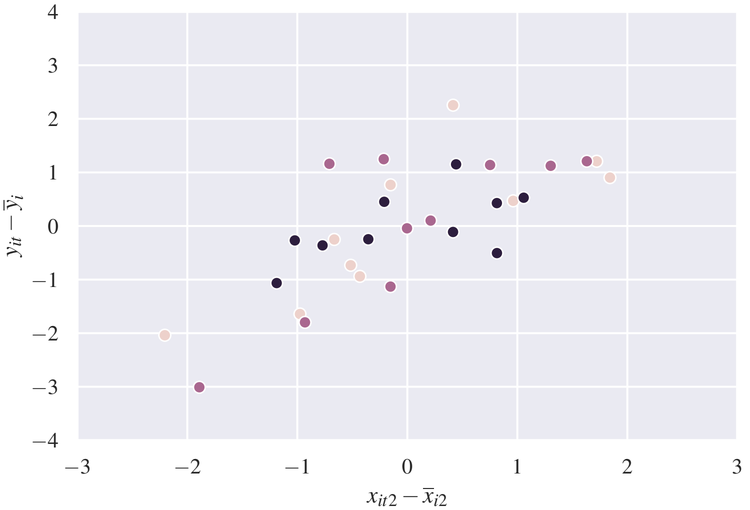
Data can be seen to ______________________

Simulated Data



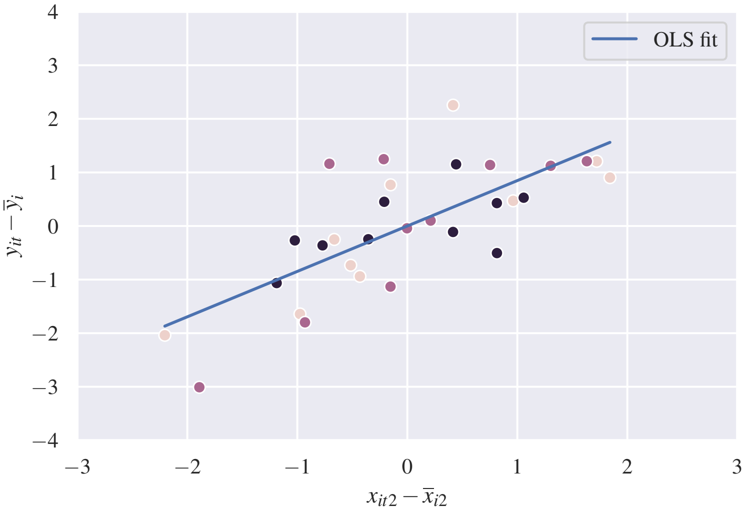
Trends?

Simulated Data



Demeaned data is ______________________________

Simulated Data



Trend in demeaned data is ____________________

Elimination or Estimation: Dummies Galore

Rewrite [BB]

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \equiv \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_i\mathbf{c} + u_{it},$$

where $\mathbf{d}_i$ gathers ID dummies

$$\mathbf{d}_i := \left[\ \mathbf{1}(i=1) \quad \mathbf{1}(i=2) \quad \cdots \quad \mathbf{1}(i=N)\ \right].$$

Could treat c_is as parameters too...

... and estimate both $\boldsymbol{\beta}$ and $\mathbf{c}$.

(Name "fixed effects" comes from this perspective.)

Elimination or Estimation: LSDV

Specifically,

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_i\mathbf{c} + u_{it} \equiv \mathbf{z}_{it}\boldsymbol{\theta} + u_{it},$$

where

$$\begin{aligned} \mathbf{z}_{it} &:= (\mathbf{x}_{it}, \mathbf{d}_i), && (1 \times (K+N)) \\ \boldsymbol{\theta} &:= (\boldsymbol{\beta}', \mathbf{c}')'. && ((K+N)\times 1) \end{aligned}$$

Least squares dummy variable (LSDV) estimator:

POLS of y_{it} on $\mathbf{x}_{it}$ and $\mathbf{d}_i$,

$$\widehat{\boldsymbol{\theta}}_{LSDV} := \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{LSDV} \\ \widehat{\mathbf{c}}_{LSDV} \end{pmatrix} := \left(\sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{Z}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{y}_i.$$

Elimination or Estimation: Discussion

Why not Estimate N Individual-Specific Intercepts?

Time-demeaning yields (algebraically!) identical $\widehat{\boldsymbol{\beta}}$

LSDVs computationally burdensome

- Inversion of $\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i$ is hard.

Consistency of POLS of $(\mathbf{c}', \boldsymbol{\beta}')'$? [D]

Identification and Consistency

Identification

Need suitable analogues of OLS.1–2 for *transformed* data.

Note: $\mathrm{E}\left[\ddot{\mathbf{x}}_{it}'\ddot{u}_{it}\right] = \mathbf{0}, t = 1, 2, \ldots, T$, follows from neither

$$\begin{aligned} & \mathrm{E}\left[\mathbf{x}_{it}u_{it}\right] = \mathbf{0}, \quad t = 1, 2, \ldots, T, \\ \underline{\text{nor}} \quad & \mathrm{E}\left[u_{it} \mid \mathbf{x}_{it}, c_i\right] = 0, \quad t = 1, 2, \ldots, T. \end{aligned}$$

Does not restrict relationship between $\mathbf{x}_{is}$ and u_{it} for $t \neq s$.

However, $\ddot{\mathbf{x}}_{it}$ and $\ddot{u}_{it}$ are based on $\mathbf{x}_{it}$s and u_{it}s for *all* t.

$\Rightarrow$ Need stronger assumption on dist'n of $\{\mathbf{x}_{it}\}_{t=1}^{T}, \{u_{it}\}_{t=1}^{T}$.

Exogeneity Condition: Strict Exogeneity

$$\textbf{FE.1 :}\ \mathrm{E}\left[u_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i\right] = 0, \quad t = 1, 2, \ldots, T.$$

In words: $\{\mathbf{x}_{it}\}_{t=1}^{T}$ are

strictly exogenous cond'l on individual (fixed) effect.

Implies $\mathrm{E}\left[\ddot{u}_{it} \mid \ddot{\mathbf{x}}_{it}\right] = 0$. [BB]

- $\ddot{\mathbf{x}}_{it}$ = fctn of $\{\mathbf{x}_{is}\}_{s=1}^{T}$ + law of iterated expectations (LIE)

Thus FE.1 $\Rightarrow$ "OLS.1" satisfied on the transformed variables.

Rank Condition

Stack $\ddot{\mathbf{x}}_{it}$s and $\ddot{y}_{it}$s over t to get $\ddot{\mathbf{X}}_i$ $(T \times K)$ and $\ddot{\mathbf{y}}_i$ $(T \times 1)$.

May write FE estimator as

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

Relevant rank condition therefore:

$$\mathbf{FE.2}: \ \operatorname{rank} \mathrm{E}[\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i] = \operatorname{rank} \left(\sum_{t=1}^{T} \mathrm{E}\left[\ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it} \right] \right) = K .$$

Conclude: FE.1–2 $\Rightarrow$ $\boldsymbol{\beta}$ identified. [BB]

Consistency

Let $\ddot{\mathbf{u}}_i$ be $T \times 1$ vector of $\ddot{u}_{it}$s. Decompose

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

LLN + Slutsky imply

$$p\text{-lim} \left(\frac{1}{N} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i \right)^{-1} \overset{\text{FE.2}}{=} \left(\mathrm{E}[\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i] \right)^{-1}.$$

Also, $N^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \ddot{\mathbf{u}}_i \xrightarrow{p} \mathrm{E}[\ddot{\mathbf{X}}_i' \ddot{\mathbf{u}}_i]$ by LLN. Now notice

$$\mathrm{E}[\ddot{\mathbf{X}}_i' \ddot{\mathbf{u}}_i] = \sum_{t=1}^{T} \mathrm{E}[\ddot{\mathbf{x}}_{it}' \ddot{u}_{it}] \overset{\text{FE.1}}{=} \mathbf{0}.$$

Conclude: FE.1–2 $\Rightarrow$ consistency of FE.

Discussion of Fixed Effects Approach

Caution: Time-constant variables in $\mathbf{x}_{it}$ net out in $\ddot{\mathbf{x}}_{it}$.

- Implies zero column in $\ddot{\mathbf{X}}_i$.
- Must then have rank $\mathrm{E}[\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i] < K$.

FE methods cannot involve e.g. ethnic origin.

- Cannot ID coefficients thereof.

Must drop such variables (and recast K).

- Why? Cannot separate effects of such var's from the c_i.

Interpretation of Strict Exogeneity

Revealing form of strict-exogeneity assumption:

$$\mathrm{E}\left[y_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i\right] = \mathrm{E}\left[y_{it} \mid \mathbf{x}_{it}, c_i\right] = \mathbf{x}_{it}\boldsymbol{\beta} + c_i \qquad (2)$$

What does it mean?

Controlling for $\mathbf{x}_{it}$, $\{\mathbf{x}_{is}, s \neq t\}$ have no partial effect on y_{it}.

Is it restrictive?

- Cannot hold if $\mathbf{x}_{it}$ contains lagged dependent variable.
- FE.1. violated for dynamic panel data models.
- Important when model captures intertemporal decisions.

Efficiency

Asymptotic Efficiency

Gather $\mathbf{x}_{it}$s in $\mathbf{x}_i := (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})$. $\qquad (1 \times KT)$

$$\textbf{FE.3}: \mathrm{E}\left[\mathbf{u}_i \mathbf{u}_i' \mid \mathbf{x}_i, c_i\right] = \sigma_u^2 \mathbf{I}_T.$$

Implies:

- (Uncond'l) Homoskedasticity, $\mathrm{E}\left[u_{it}^2\right] = \sigma_u^2$ all t.
- (Uncond'l) No serial correlation, $\mathrm{E}\left[u_{it} u_{is}\right] = 0$ all $t \neq s$.

In fact, stronger:

- Assumes *conditional* (on $\mathbf{x}_i, c_i$) variances constant
- ... and *conditional* covariances zero.

Asymptotic Efficiency

Using FE.1–2, some work (do it!) shows

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

$$\text{where} \quad \mathbf{A} := \mathrm{E}[\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i] \quad \text{and} \quad \mathbf{B} := \mathrm{E}[\ddot{\mathbf{X}}_i'\mathbf{u}_i\mathbf{u}_i'\ddot{\mathbf{X}}_i].$$

Under FE.3,

$$\begin{aligned} \mathrm{E}[\ddot{\mathbf{X}}_i'\mathbf{u}_i\mathbf{u}_i'\ddot{\mathbf{X}}_i] &= \mathrm{E}[\ddot{\mathbf{X}}_i'\mathrm{E}\left(\mathbf{u}_i\mathbf{u}_i'|\mathbf{x}_i, c_i\right)\ddot{\mathbf{X}}_i] && \text{(LIE)} \\ &= \mathrm{E}\left[\ddot{\mathbf{X}}_i'\sigma_u^2\mathbf{I}_T\ddot{\mathbf{X}}_i\right]. && \text{(FE.3)} \end{aligned}$$

So $\mathbf{B} = \sigma_u^2\mathbf{A}$ and, thus,

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{FE}) = \sigma_u^2\left(\mathrm{E}[\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i]\right)^{-1}/N.$$

Conclude: FE.1–3 $\Rightarrow$ FE asymptotically efficient.

Inference

Standard Variance Estimation

Under FE.1–3, consistent estimator:

$$\widehat{\mathrm{Avar}(\widehat{\beta}_{FE})} := \widehat{\sigma}_u^2 \left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right)^{-1},$$

$$\text{where} \quad \widehat{\sigma}_u^2 := \frac{1}{NT - N - K} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\ddot{u}}_{it}^2,$$

$$\widehat{\ddot{u}}_{it} := \ddot{y}_{it} - \ddot{\mathbf{x}}_{it} \widehat{\beta}_{FE}. \qquad \text{(FE residuals)}$$

"Standard" OLS variance estimator for transformed data.

Well, almost...

Potential Pitfall

$$\widehat{\sigma}_u^2 = \frac{1}{NT - N - K} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\ddot{u}}_{it}^2,$$

Caution: Fewer degrees of freedom, $NT - (N + K)$.

- Implicit estimation of individual intercepts. (LSDV...)

Without $-N$, resulting estimator inconsistent.

- Not just a d.f. "correction." Matters greatly.

  In practice: Multiply "default" s.e.'s by

$$\sqrt{(NT - K) / (NT - N - K)}.$$

Robust Variance Estimation

FE.3 rules out (cond'l) homoskedasticity and serial correlation.

May be violated/hard to motivate in empirical setting.

Robust variance estimator given by

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

$$\text{where} \quad \widehat{\ddot{\mathbf{u}}}_i := (\widehat{\ddot{u}}_{i1}, \ldots, \widehat{\ddot{u}}_{iT})'. \qquad (T \times 1)$$

Consistent under FE.1–2 alone.

Inference possible w/o FE.3. (= my recommendation.)

First-Difference Methods

First-Difference Transformation and Estimation

First-Difference (FD) transformation

Similar progression albeit w/ new transformation.

Lack and diff' to get

$$
\begin{aligned}
y_{it} - y_{it-1} &= (\mathbf{x}_{it} - \mathbf{x}_{it-1})\boldsymbol{\beta} + \overbrace{(c_i - c_i)}^{=0} + u_{it} - u_{it-1} \\
\Leftrightarrow \Delta y_{it} &= \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}, \quad t = \textcolor{red}{2}, 3, \ldots, T.
\end{aligned}
$$

Also eliminates time-constant c_i.

One time period lost due to diff'ing

Interpretation of $\boldsymbol{\beta}$ comes from structural cond'l expectation.

Differenced Data and OLS

FD estimator = pooled OLS after differencing

$$\hat{\beta}_{FD} := \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\mathbf{x}_{it}'\Delta\mathbf{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\mathbf{x}_{it}'\Delta y_{it}\right)$$
$$= \left(\sum_{i=1}^{N}\Delta\mathbf{X}_i'\Delta\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\Delta\mathbf{X}_i'\Delta\mathbf{y}_i\right)$$
$$= (\Delta\mathbf{X}'\Delta\mathbf{X})^{-1}\,\Delta\mathbf{X}'\Delta\mathbf{y}$$

where

- $\Delta\mathbf{X}_i$ is $(T-1)\times K$, (when all time-varying)
- $\Delta\mathbf{y}_i$ is $(T-1)\times 1$, and matrices stack over i

Caution: Beware of ghost observations. [BB]

Identification and Consistency

Identification and Consistency

Relevant exogeneity condition:

$$\mathbf{FD.1}(=\mathbf{FE.1}): \ \mathrm{E}\left[u_{it} \mid \mathbf{x}_i, c_i\right] = 0, \quad t = 1, 2, \ldots, T.$$

$$\Rightarrow \mathrm{E}\left[\Delta \mathbf{x}_{it}' \Delta u_{it}\right] = 0, \quad t = 2, 3, \ldots, T.$$

- I.e. "OLS.1" satisfied after FD'ing.

Actually, $\mathrm{E}[u_{it} \mid \mathbf{x}_{it-1}, \mathbf{x}_{it}, \mathbf{x}_{it+1}] = 0$ suffices, since

$$\begin{aligned}\mathrm{E}\left[\Delta \mathbf{x}_{it}' \Delta u_{it}\right] &= \mathrm{E}\left[\mathbf{x}_{it}' u_{it}\right] - \mathrm{E}\left[\mathbf{x}_{it-1}' u_{it}\right] \\ &\quad - \mathrm{E}\left[\mathbf{x}_{it}' u_{it-1}\right] + \mathrm{E}\left[\mathbf{x}_{it-1}' u_{it-1}\right].\end{aligned}$$

Identification and Consistency

Rewrite

$$\widehat{\boldsymbol{\beta}}_{FD} = \left(\frac{1}{N} \sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{y}_i \right).$$

Relevant rank condition then

$$\textbf{FD.2}: \ \operatorname{rank} \mathrm{E}\left[\Delta \mathbf{X}_i' \Delta \mathbf{X}_i\right] = \operatorname{rank} \left(\sum_{t=2}^{T} \mathrm{E}\left[\Delta \mathbf{x}_{it}' \Delta \mathbf{x}_{it}\right] \right) = K.$$

- (Still) Cannot ID/estimate effects of time-constant var's.

Under FD.1–2, $\boldsymbol{\beta}$ identified + $\widehat{\boldsymbol{\beta}}_{FD}$ consistent. (Check!)

Efficiency

Asymptotic Efficiency of First Differences

For efficiency of FD estimator, suffices that also

$$\textbf{FD.3:}\ \mathrm{E}\left[\mathbf{e}_i\mathbf{e}_i' \mid \mathbf{x}_i, c_i\right] = \sigma_e^2 \mathbf{I}_{T-1}, \qquad ((T-1)\times(T-1))$$

where $e_{it} := \Delta u_{it}$ for $t = 2, 3, ..., T$.

FD.1–3 $\Rightarrow$ FD asymptotically efficient.

[Check!]

Efficiency: Discussion

$$\textbf{FD.3: } \mathrm{E}\left[\mathbf{e}_i\mathbf{e}_i' \mid \mathbf{x}_i, c_i\right] = \sigma_e^2 \mathbf{I}_{T-1}, \qquad ((T-1)\times(T-1))$$

- Implies $\{\Delta u_{it}\}_{t=2}^{T}$ homoskedastic + serially uncorrelated.
- Under FD.3, $u_{it} = u_{it-1} + e_{it}$—random walk.
- Conversely, if u_{it}s initially uncorrelated (FE.3), FD'ing induces serial correlation. [Check!]
- FD.3 and FE.3 represent opposite extremes.

Inference

Variance Estimation

Under FD.1–3, consistent estimator

$$\widehat{\text{Avar}(\hat{\beta}_{FD})} := \widehat{\sigma}_e^2 \left(\Delta \mathbf{X}' \Delta \mathbf{X}\right)^{-1},$$

$$\text{where} \quad \widehat{\sigma}_e^2 := \frac{1}{N(T-1)-K} \sum_{i=1}^{N} \sum_{t=2}^{N} \widehat{e}_{it}^2,$$

$$\widehat{e}_{it} := \widehat{\Delta u_{it}} := \Delta y_{it} - \Delta \mathbf{x}_{it} \widehat{\beta}_{FD}. \qquad \text{(FD residuals)}$$

- Equal to "standard" OLS variance estimator.
- $N(T-1)$ = effective sample size (due to differencing).

Robust Variance Estimation

May not be willing to impose FD.3.

Under FD.1–2 alone (check!),

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{FD} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right),$$

$$\text{where} \quad \mathbf{A} := \mathrm{E}\left[\Delta\mathbf{X}_i'\Delta\mathbf{X}_i\right] \quad \text{and} \quad \mathbf{B} := \mathrm{E}\left[\Delta\mathbf{X}_i'\mathbf{e}_i\mathbf{e}_i'\Delta\mathbf{X}_i\right].$$

Robust variance estimator is

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{FD})} := \left(\Delta\mathbf{X}'\Delta\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{N}\Delta\mathbf{X}_i'\widehat{\mathbf{e}}_i\widehat{\mathbf{e}}_i'\Delta\mathbf{X}_i\right)\left(\Delta\mathbf{X}'\Delta\mathbf{X}\right)^{-1}.$$

Consistent (hence inference possible) under FD.1–2 alone.

To FE or not to FE (i.e. FD)

- Under FE/D.1, choice FE/D about efficiency (2^{nd} order).
- When $T = 2$, FD and FE algebraically equivalent.
- Follows from (check!)

$$\ddot{y}_{i2} = y_{i2} - \frac{y_{i1} + y_{i2}}{2} = \frac{1}{2}\Delta y_{i2},$$
$$\ddot{\mathbf{x}}_{i2} = \frac{1}{2}\Delta \mathbf{x}_{i2},$$
$$\ddot{y}_{i1} = -\frac{1}{2}\Delta y_{i2},$$
$$\ddot{\mathbf{x}}_{i1} = -\frac{1}{2}\Delta \mathbf{x}_{i2}.$$

- If strict exog. unreasonable, change strategy. (Later.)