Causal Panel Designs

BUSS975: Causal Inference in Financial Research

Professor Ji-Woong Chung

Korea University

Where we left off

- Lecture 5 spent its time looking for a **shock**: a lottery, a judge, a national surge, a threshold, the weather. Something that moved treatment for reasons the unit did not choose.
- Good instruments are scarce. Most of the time you do not have one.

- Panel data offers a different trade. Instead of finding a shock, **use the unit as its own control**.
- If you see the same firm in 2015 and in 2019, anything about that firm which did not change in between cannot explain why its outcome changed.

The plan

1. **What panel data is** — and what the structure buys you.
2. **Pooled OLS**, and the two separate ways it fails.
3. **Fixed effects**, the *within* estimator: the transformation, three ways to compute it, the assumptions.
4. **Three estimators side by side**, on data where we know the right answer.
5. **Inference**: clustering, and what to do when clusters are few.
6. **What fixed effects cannot fix** — two identification failures, then four costs.
7. **Fixed effects in finance**: richer effects, and the comparison each one leaves.

1. What Panel Data Is

Two things that get called the same name

	You follow	Example
Panel (longitudinal)	the *same* units over time	Compustat firm-years
Repeated cross-section	a fresh sample each period	the Census, most surveys

- Only a true panel lets you difference a unit against **itself**. A repeated cross-section lets you difference *groups* against themselves, which is a weaker but still useful thing.
- Notation throughout: $i = 1, \ldots, N$ units, $t = 1, \ldots, T$ periods. We assume a **balanced** panel, so every unit appears in every period.

Motivation: a regression you have all run

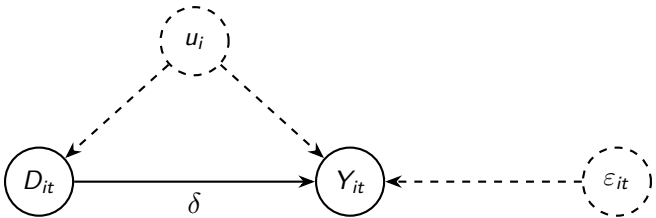
$$\text{leverage}_{i,j,t} = \beta_0 + \beta_1 \text{profitability}_{i,j,t} + u_{i,j,t}$$

for firm i in industry j in year t. **What is in u?**

- Managerial talent and managerial risk aversion. Industry supply and demand shocks. The firm's cost of capital and its investment opportunity set. Local institutions and financial development.

- Every one of these plausibly moves leverage **and** correlates with profitability.
- **Why not just use a proxy?** A proxy delivers consistency only if the unobservable is uncorrelated with the regressors *after* conditioning on it. That is a demanding requirement, and you cannot check it.

The confounding, as a picture



- u_i is a **confounder**: it causes the treatment and it causes the outcome.
- Ordinarily that would be the end of the story — you cannot condition on what you cannot see.
- **But u_i carries no t subscript.** It is the same number in every period for a given unit.

The model

For each unit i in period t:

$$Y_{it} = \alpha + \delta D_{it} + \underbrace{u_i}_{\substack{\text{unobserved} \\ \text{heterogeneity}}} + \underbrace{\varepsilon_{it}}_{\substack{\text{idiosyncratic} \\ \text{error}}}, \qquad t = 1, \ldots, T.$$

Reading it with the schooling–earnings example:

Y_{it}	log earnings of person i in year t
D_{it}	schooling — the treatment, and it must **change** over time
δ	the returns to schooling — assumed the *same* for everyone
u_i	everything about i that never changes: ability, upbringing, drive
ε_{it}	everything that changes period to period: luck, health, the cycle

2. Pooled OLS

Ignoring the panel structure

Throw the time dimension away and regress Y_{it} on D_{it}:

$$Y_{it} = \alpha + \delta D_{it} + \eta_{it}, \qquad \eta_{it} \equiv u_i + \varepsilon_{it}.$$

- This is **pooled OLS (POLS)**.
- POLS needs only **contemporaneous exogeneity**: the error is unrelated to treatment *in the same period*,

$$E[\eta_{it} \mid D_{it}] = 0, \qquad t = 1, \ldots, T.$$

- But u_i sits inside η_{it}, so this requires u_i to be **uncorrelated with treatment** — exactly what the picture denies.

POLS fails twice, not once

- **Failure 1: bias.** u_i is in the error and correlated with D_{it}. $\Rightarrow$ Omitted variable bias.
- **Failure 2: the standard errors.** u_i appears in η_{it} for **every** period of unit i. So the errors of the same unit are correlated with each other by construction — serial correlation.
  - The reported standard errors are therefore **too small**, often by a lot. You get a tight interval around a biased number.

3. Fixed Effects: The Within Estimator

The idea in one line

Stop comparing person i to person j. Compare person i to ***person*** *i* ***at another date****.*

- Any confounder that is the same in both of i's periods cannot explain a change between them.
- This works **whether or not you can observe the confounder**, and whether or not you can name it.
- The price is equally sharp: the only variation left is variation **within** units. Units that never change contribute nothing.
- So the estimate is built entirely from **switchers** — people whose schooling changed, firms whose policy changed. Name them before you interpret the number.

Where did u_i go?

Average the model over time for a single unit i:

$$\bar{Y}_i = \alpha + \delta \bar{D}_i + u_i + \bar{\varepsilon}_i, \qquad \text{where } \bar{Y}_i = \tfrac{1}{T} \textstyle\sum_t Y_{it}.$$

Now subtract that from the original equation, period by period:

$$\underbrace{(Y_{it} - \bar{Y}_i)}_{\ddot{Y}_{it}} = \delta \underbrace{(D_{it} - \bar{D}_i)}_{\ddot{D}_{it}} + \underbrace{(\alpha - \alpha)}_{=\,0} + \underbrace{(u_i - u_i)}_{=\,0} + \underbrace{(\varepsilon_{it} - \bar{\varepsilon}_i)}_{\ddot{\varepsilon}_{it}}.$$

$$\boxed{\ddot{Y}_{it} = \delta \ddot{D}_{it} + \ddot{\varepsilon}_{it}}$$

- The price: $\ddot{\varepsilon}_{it}$ contains $\bar{\varepsilon}_i$, which averages **every** period's shock. So D_{it} must be unrelated to the shocks of *all* periods, not just its own — **strict exogeneity**.

The same thing, as a minimisation problem

Treat the u_i as N parameters to be estimated alongside δ:

$$(\hat{\delta},\ \hat{u}_1, \ldots, \hat{u}_N) \ = \ \underset{b,\, m_1, \ldots, m_N}{\arg\min} \ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - D_{it} b - m_i\right)^2.$$

- This is a regression of Y_{it} on D_{it} **plus** N **individual dummies**.
- The first-order condition with respect to m_i gives, for each unit,

$$\hat{m}_i \ = \ \frac{1}{T} \sum_{t=1}^{T} \left(Y_{it} - D_{it} \hat{\delta}\right) \ = \ \bar{Y}_i - \bar{D}_i \hat{\delta}.$$

- The estimated fixed effect $\hat{m}_i$ is simply unit i's average residual.

Why the two routes coincide

The first-order condition with respect to b is:

$$\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}\left(Y_{it} - D_{it}\hat{\delta} - \hat{m}_i\right) = 0.$$

Substituting $\hat{m}_i = \bar{Y}_i - \bar{D}_i\hat{\delta}$ back into this equation:

$$\hat{\delta} \;=\; \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(D_{it} - \bar{D}_i)^2\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(D_{it} - \bar{D}_i)(Y_{it} - \bar{Y}_i)\right)$$

$$\hat{\delta} \;=\; \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{D}_{it}^2\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{D}_{it}\ddot{Y}_{it}\right)$$

Three ways to compute the same number

	What you do	Watch out for
1. Demean	regress $\ddot{Y}_{it}$ on $\ddot{D}_{it}$ by hand	degrees of freedom
2. Unit dummies	regress Y_{it} on D_{it} and N dummies	large N
3. Canned routine	`feols`, `plm`, `xtreg, fe`	—

- All three give the **numerically identical** slope.
- The **within** estimator, the **fixed effects** estimator, the **dummy variable** regression (LSDV), the **demeaned** regression. Add time dummies as well and it becomes **two-way fixed effects**.

What fixed effects silently deletes

- Demeaning removes u_i. It removes **every** time-invariant variable the same way, including ones you wanted to keep.
- If your primary parameter of interest is fixed over time, standard FE **cannot** estimate it.

- Variables that change rarely (e.g., institutional quality, policy regimes) retain almost no within-unit variation after demeaning.
- Estimates become imprecise with inflated standard errors.

- **Workarounds:**
  - Interact X_i with time indicators ($X_i \times \text{Year}_t$) to see how its effect evolves over time.
  - Hausman–Taylor (1981): use time-varying exogenous regressors as internal instruments for time-invariant ones.

A trap you can actually fall into

Estimate $\ln(\text{pay})_{ijt}$ with CEO, firm and year effects, and include $\texttt{female}_i$. The CEO effects already absorb it. What does the software do?

- `xtreg`-style routines drop `female` and say so. **Fine.**
- But build the dummies yourself and put them in the regression, and the software may drop **one of your dummies instead** — then report a coefficient on `female`.
- That number is not an estimate of anything. It is an artifact of which column the solver happened to discard.

- The lesson: if a time-invariant regressor survives a fixed-effects regression, something is wrong — find out what got dropped in its place.

Strict exogeneity is stronger than it looks

$$E[\varepsilon_{it} \mid D_{i1}, D_{i2}, \ldots, D_{iT},\ u_i] = 0, \qquad t = 1, \ldots, T.$$

- POLS asked only that D_{it} be unrelated to *this* period's shock. FE asks that it be unrelated to the shock of **every** period.
- The conditioning set includes **future** treatment, so strict exogeneity rules out **feedback**: a shock today cannot influence treatment tomorrow.
- Say Y is profitability and D is investment. A firm has a bad year ($\varepsilon_{it} < 0$) and cuts investment next year. Then ε_{it} predicts $D_{i,t+1}$, and strict exogeneity fails.
- **This is the assumption that fails most often in finance**, because firms react to what happens to them.

A lagged dependent variable violates it too — *by construction*. That is the **dynamic panel** problem.

It was never really about time

- Fixed effects works wherever observations can be **grouped**, removing any unobserved heterogeneity that is constant *within a group*.
- Firms grouped by **industry**. Patents grouped by **examiner**. Loans grouped by **bank–year**. CEO spells grouped by **CEO–firm** pairs. Students grouped by **classroom**.
- The logic is the same every time: compare units **inside** a group, so anything the group shares cancels.

4. Three Estimators, Side by Side

A firm panel we can check against the truth

Simulate a panel where we **set** the answer.

i, t	300 firms, 5 years each
u_i	managerial risk appetite — *unobserved*, fixed per firm
$\texttt{prof}_{it}$	profitability, built to be correlated with u_i
$\texttt{founder}_i$	founder-led — a *time-invariant* trait, also correlated with u_i
$\texttt{lev}_{it}$	leverage: $0.5 - 0.30\,\texttt{prof}_{it} + 0.20\,\texttt{founder}_i + u_i + \varepsilon_{it}$

- **The truth is** -0.30: more profitable firms borrow less.
- u_i raises leverage *and* is correlated with profitability and with founder status

Three estimators, one number

Dependent variable: leverage	POLS	LSDV / FE	Demeaned OLS
Profitability (truth -0.30)	0.057	-0.352	-0.352
	(0.037)	(0.030)	(0.030)
Founder-led (truth $+0.20$)	0.834	—	—
	(0.108)		
Firm dummies	no	yes	absorbed
R^2	0.100	0.523	0.104

Standard errors clustered by firm, computed as `xtreg, fe` and `reghdfe` do: firm effects nested in the clusters are not counted as parameters. $N = 1{,}500$ firm-years, 300 firms.

- In Stata, `xtreg, fe` reports the within R^2; `areg` reports an overall R^2 that includes what the fixed effects explain. Know which one you are quoting.

5. Inference

Cluster by unit

- **Keep two decisions separate.**
  - **Fixed effects** fix the *point estimate*: they sweep out group means that would bias $\hat{\delta}$.
  - **Clustering** fixes the *standard error*: it lets shocks be correlated within a group.
- Within a unit, the ε_{it} are almost certainly correlated over time. **Cluster by panel unit** — or, if treatment is assigned to whole groups of units (a state policy studied on worker data), by that group.

- Cluster-robust inference is valid as the **number of clusters** G grows, not N.
- Rule of thumb: < 40 clusters, treat the standard errors with suspicion. Remedies: the **wild cluster bootstrap** (Cameron, Gelbach and Miller 2008), or small-sample corrections (Bell and McCaffrey 2002).

The few-clusters problem

- Cluster-robust standard errors (**CRSE**) are built from **one score per cluster**. With $G = 10$ you are estimating a variance from ten numbers.
- With few clusters (fewer than about 40), CRSE **understates** sampling variability. The t-statistics are too big, and a 5% test rejects a true null too often.

- What is usually scarce is not clusters but **treated** clusters. If 2 of 10 states are treated, $\hat{\delta}$ is driven by those two states' shocks.
- The treatment dummy soaks up much of those shocks, so the treated states' residuals look small — and the CRSE looks precise when it is not.
- Unequal cluster sizes make it worse (MacKinnon and Webb 2017).

The wild cluster bootstrap: intuition

Model: $Y_{it} = \delta D_{it} + X_{it}\beta + \alpha_i + \lambda_t + \varepsilon_{it}$, with clusters $g = 1, \ldots, G$; $g(i)$ is unit i's cluster.

- Do not resample observations or clusters. **Keep every regressor fixed and resample the residuals** — one coin per cluster.

1. Fit the model **imposing** $H_0 : \delta = 0$. Keep its fitted values $\tilde{Y}_{it}$ and residuals $\tilde{\varepsilon}_{it}$.
2. For each cluster, toss a coin. Heads: keep its residuals. Tails: flip the sign of **all** of them.

$$Y^*_{it} \;=\; \tilde{Y}_{it} + w_{g(i)}\,\tilde{\varepsilon}_{it}, \qquad w_g \in \{-1, +1\}.$$

- **Why it works:** flipping a whole cluster at once keeps its within-cluster correlation intact, and the pseudo-data satisfy H_0 by construction.

The wild cluster bootstrap: algorithm

1. **Restricted fit.** Estimate the model with $\delta = 0$ imposed. Keep $\tilde{Y}_{it}$ and $\tilde{\varepsilon}_{it}$.
2. **Pseudo-outcomes.** For $b = 1, \ldots, B$: draw a weight w_g for each cluster and set $Y_{it}^{(b)} = \tilde{Y}_{it} + w_{g(i)}\,\tilde{\varepsilon}_{it}$.
3. **Re-estimate.** Regress $Y_{it}^{(b)}$ on D_{it}, X_{it} and the fixed effects; record the cluster-robust t-statistic $t^{(b)}$ for $\hat{\delta}$.
4. **p-value.** Locate the actual t in the bootstrap distribution:

$$p_{\text{WCB}} \;=\; \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\big(|t^{(b)}| \geq |t|\big).$$

- With G clusters there are only 2^G sign patterns (1,024 for $G = 10$). Below $G \approx 12$, use Webb's six-point weights instead (Webb 2023).

Numerical example: building the panel

```
clear all
set seed 12345
local n_treated 5                        // number of treated states; try 2

* 1. 10 states x 100 workers; worker effect drawn ONCE per worker
set obs 1000
gen worker_id = _n
gen state = ceil(worker_id / 100)
gen worker_effect = rnormal(15, 2)

* 2. Expand across 5 years (2020-2024): 5,000 observations
expand 5
bysort worker_id: gen year = 2020 + _n - 1

* 3. The policy starts in 2022 in states 1, ..., n_treated
gen policy_subsidy = (state <= `n_treated' & year >= 2022)
```

- 5,000 observations — but only **ten clusters**, and we choose how many are treated.
- `worker_effect` is drawn *before* `expand`, so each worker keeps one value in every year.

Numerical example: wages with no true effect

```
* 4. One shock per state-year, shared by every worker in it
bysort state year: gen state_year_shock = rnormal(0, 1.2) if _n == 1
by state year: replace state_year_shock = state_year_shock[1]

* 5. Wages with ZERO true policy effect
sort worker_id year
gen wage = worker_effect + state_year_shock + rnormal(0, 1)
xtset worker_id year
```

- **The true effect is exactly zero.** Any rejection is a false positive.
- The state-year shock correlates residuals within a state — what clustering exists to handle.
- **A Stata trap:** `bysort state year: gen x = rnormal()` draws a new number for *every row*, not one per group. The `_n == 1` / `[1]` pair is what makes the shock common.

Stata implementation

```
cap ssc install ftools
cap ssc install reghdfe
cap ssc install boottest

* A. Standard CRSE, clustered by state (G = 10). Year enters as dummies:
*    boottest allows at most one absorbed fixed effect after reghdfe
reghdfe wage policy_subsidy i.year, absorb(worker_id) vce(cluster state)

* B. WCB with Rademacher signs. Only 2^10 = 1,024 sign patterns exist;
*    asking for more reps makes boottest enumerate all of them
boottest policy_subsidy, weight(rademacher) reps(9999) seed(12345)

* C. WCB with Webb's six-point weights (6^10 patterns): preferred below G = 12
boottest policy_subsidy, weight(webb) reps(9999) seed(12345)
```

- `boottest` reads the model and the clustering from the preceding command (`regress`, `areg`, `xtreg`, `ivreg2`, or `reghdfe` with one absorbed effect) and inverts the test into a confidence interval (Roodman et al. 2019).
- In R: `fwildclusterboot::boottest()` — same algorithm.

What happens across 5,000 such panels

$\text{Wage}_{it} = \delta\,\text{Policy}_{it} + \alpha_i + \lambda_t + \varepsilon_{it}$, true $\delta = 0$, $G = 10$ states. **How often does a 5% test reject?**

Treated states (of 10)	Standard CRSE	WCB (Webb weights)
2	17.1%	0.6%
3	11.0%	7.0%
5	6.3%	5.7%

Share of 5,000 simulated panels (the DGP above) that reject $H_0 : \delta = 0$ at 5%. CRSE uses t_{G-1} critical values, as `reghdfe` does; WCB is restricted, $B = 999$. Code: `06a_panel_wcb_sim.R`.

- **Few *treated* clusters is the danger.** With 2 treated states, CRSE rejects a true null 17% of the time.
- **WCB does not rescue that case.** It almost never rejects, so it has almost no power either (MacKinnon and Webb 2017). No procedure can manufacture variation that two treated states do not contain.
- **With more treated clusters, both move toward** 5%, and WCB gets there faster. Count your *treated* clusters before trusting any p-value.

6. What Fixed Effects Cannot Fix

Reverse causality

Becker (1968) predicts that more policing lowers crime. Cornwell and Trumbull (1994):

DepVar: crime rate	Between	Within	2SLS (FE)	2SLS (no FE)
Police	0.364	0.413	0.504	0.419
	(0.060)	(0.027)	(0.617)	(0.218)

County-level panel. Standard errors in parentheses. All columns include controls.

- **Between** regresses county *averages* on county averages: cross-county variation only, the complement of within.
- Every column says more police, **more** crime — and fixed effects make it *larger*.
- The reason: **Counties hire police because crime is high**: $Y \to D$. Demeaning cannot touch that.

Time-varying unobserved heterogeneity

Married men earn more. Is that marriage, or is it the kind of man who marries? Cornwell and Rupert (1997):

DepVar: log wage	RE (FGLS)	Within	Within	Within
Married	0.083	0.056	0.051	0.033
	(0.022)	(0.026)	(0.026)	(0.028)
Education controls	Yes	No	No	No
Tenure	No	No	Yes	Yes
Quadratics in years married	No	No	No	Yes

- **FE removes the fixed part of selection** (ability, temperament): $0.083 \rightarrow 0.056$.
- **Time-varying controls move it again**, to an insignificant 0.033. Something that *changes over time* is tangled up with marriage, beyond the reach of FE.
- **Confounder or channel?** If marriage raises tenure, controlling for tenure removes part of the effect. The DAG decides, not FE.

Cost 1: measurement error gets worse

You observe $x = x^* + e$: **signal** plus **noise**. In levels, classical attenuation says

$$\operatorname{plim} \hat{\delta} \;=\; \delta \cdot \frac{\operatorname{Var}(x^*)}{\operatorname{Var}(x^*) + \operatorname{Var}(e)} \;<\; \delta.$$

- Fixed effects removes the **cross-sectional** part of the variation. For most regressors that is where the signal lives — firms differ from one another more than any one firm differs from itself.
- The noise is idiosyncratic, so it largely survives the demeaning.
- The ratio therefore **falls**, and attenuation gets *worse*.

- Insignificance may reflect noise attenuation, not a true zero effect.
- Mismeasurement leaks bias into other control coefficients in unknown directions.

But it depends on what persists

Write ρ for the autocorrelation of the **signal** and r for the autocorrelation of the **noise**. For first differences (FE behaves the same way; the algebra is messier):

$$\operatorname{plim} \hat{\delta} \;=\; \delta \cdot \frac{1}{1 + \dfrac{\sigma_e^2 (1 - r)}{\sigma_{x^*}^2 (1 - \rho)}}.$$

Case	Meaning	Consequence
$r = \rho = 0$	both signal and noise iid	ordinary attenuation
$r = 0,\ \rho$ high	persistent signal, transitory noise	**much worse**
$r = 1$	the mismeasurement is *constant*	FE **eliminates** the error

- Fixed effects worsens measurement error **whenever the noise is less persistent than the signal**.

Aside: where that formula comes from

Classical measurement error: $x_{it} = x_{it}^* + e_{it}$, with e uncorrelated with x^* and with the equation error. Differencing and applying the same attenuation logic,

$$\operatorname{plim} \hat{\delta}_{\mathrm{FD}} \;=\; \delta \cdot \frac{\operatorname{Var}(\Delta x^*)}{\operatorname{Var}(\Delta x^*) + \operatorname{Var}(\Delta e)}.$$

If x^* and e are **stationary**, with first-order autocorrelations ρ and r,

$$\operatorname{Var}(\Delta x^*) = 2\sigma_{x^*}^2 - 2\operatorname{Cov}(x_t^*, x_{t-1}^*) = 2\sigma_{x^*}^2(1-\rho), \qquad \operatorname{Var}(\Delta e) = 2\sigma_e^2(1-r).$$

Substitute; the 2's cancel:

$$\operatorname{plim} \hat{\delta}_{\mathrm{FD}} = \delta \cdot \frac{2\sigma_{x^*}^2(1-\rho)}{2\sigma_{x^*}^2(1-\rho) + 2\sigma_e^2(1-r)} = \frac{\delta}{1 + \dfrac{\sigma_e^2(1-r)}{\sigma_{x^*}^2(1-\rho)}}.$$

Cost 2: the same thing, without any mismeasurement

- Suppose Y responds to **sustained** changes in X, not to transitory wobbles — investment responds to a permanent shift in the opportunity set, not to one noisy quarter.
- Fixed effects throws away the permanent component by construction. What is left in $\ddot{X}_{it}$ is largely the transitory wobble.
- You can then find X unrelated to Y under fixed effects, **even though sustained changes in X are the main determinant of Y.**

- Fixed effects answers a question about within-unit variation; if your theory is about between-unit differences, you have changed the question.

Cost 3: you cannot trust the $\hat{u}_i$ themselves

Sometimes the fixed effect *is* the object of interest. Recover it as $\hat{u}_i = \bar{Y}_i - \bar{D}_i\hat{\delta}$ — Bertrand and Schoar (2003) did exactly this to back out **CEO effects**, and argued that managers have persistent "styles."

- **But $\hat{u}_i$ is unbiased and *inconsistent*.** As the sample grows you get more groups, so more parameters — never more data *per* parameter. The asymptotics you need never arrive. This is the **incidental parameters problem**.
- Note the asymmetry: $\hat{\delta}$ is fine as $N \to \infty$ with T fixed. The $\hat{u}_i$ are not.
- And the F-test that the $\hat{u}_i$ are jointly non-zero is valid only under strong assumptions — errors normal, homoskedastic, serially uncorrelated.
- Fee, Hadlock and Pierce (2013) press exactly that point: apparent manager "style" survives procedures it should not, so the F-test is not evidence of style.

Cost 4: and none of this works in a nonlinear model

- **You cannot demean to eliminate u_i in a nonlinear model.** Probit, Logit and Tobit have no within transformation, so the u_i must be estimated, and their inconsistency contaminates the coefficient you actually wanted.
- The escape routes demand assumptions most people would not sign:
  - **Probit:** estimating the u_i as dummies is inconsistent for fixed T (Greene 2004). The usual fallback, *random-effects* probit, needs u_i normal *and independent of* D_{it} — the very correlation fixed effects existed to permit.
  - **Logit:** conditional logit works, but needs no serial correlation in Y given D and u_i, and discards every unit whose outcome never changes.

- **Rule of thumb:** Avoid high-dimensional FE in nonlinear settings. Prefer Linear Probability Models (LPM) or **Poisson FE** (which avoids incidental parameter bias).

7. Fixed Effects in Finance

What firm and year effects absorb

The workhorse specification in corporate finance:

$$Y_{it} = \delta D_{it} + X_{it}\beta + \underbrace{\alpha_i}_{\text{firm}} + \underbrace{\lambda_t}_{\text{year}} + \varepsilon_{it}.$$

- **Firm effects** α_i absorb anything fixed about the firm: its industry, location, culture.
- **Year effects** λ_t absorb anything common to all firms in a year: the cycle, interest rates, a new regulation.
- What is left is **firm-year** variation: how far *this* firm moved in *this* year, beyond its own usual level and beyond what every firm did that year.

Two-way FE by hand: two firms, two years

Truth: $Y_{it} = 2D_{it} + \alpha_i + \lambda_t$, no noise. Firm A is big ($\alpha_A = 10$), B is small ($\alpha_B = 4$), and year 2 is a boom ($\lambda_2 = 3$). Only A is treated, only in year 2.

	Year 1	Year 2	Change
A $(D = 0, 1)$	10	15	+5
B $(D = 0, 0)$	4	7	+3

Estimator	$\hat{\delta}$	What contaminates it
Pooled OLS	8	A is big, *and* year 2 booms
Firm FE only	5	the boom
Year FE only	8	A is big
Firm + year FE	**2**	nothing

- Each effect removes one confounder: firm effects remove *size*, year effects remove the *boom*. You need both.
- $\hat{\delta} = (15 - 10) - (7 - 4) = 2$: on a 2×2 panel, two-way FE **is** difference-in-differences.

Richer fixed effects, and what each one compares

- Finance keeps asking for effects where **both** dimensions are large. Naive dummies would mean inverting a matrix with tens of thousands of columns; `reghdfe` (Stata) and `fixest::feols` (R) made it routine.

Fixed effects	The second effect absorbs	What you are left comparing
firm + industry×year	the industry's whole year: demand, input prices, industry regulation	a firm against *its own industry peers*, in the same year
bank + county×quarter	local credit *demand* and local conditions	two banks lending into the *same market, same quarter*
CEO + firm + year	the firm permanently, and the macro year	the *same CEO at a different firm* — movers only

- Each added effect absorbs a confounder you **cannot measure** but which has a *group* × *time* shape.
- The price is always the same: any treatment that varies **only** at that level is absorbed too.

Firm + industry × year, in numbers

Truth: $Y_{it} = 2D_{it} + \alpha_i + \lambda_t + \eta_{jt}$. Every firm gets $\lambda_2 = 1$; **industry 1 alone** gets a shock $\eta_{12} = 3$. Only A is treated, in year 2.

Industry	Firm		Year 1	Year 2	Change
1	A	$(D = 0, 1)$	10	16	+6
1	B		6	10	+4
2	C		8	9	+1
2	D		4	5	+1

- **Firm + year:** A's change against the *average* of the other three: $6 - \frac{4+1+1}{3} = 4$. Two-thirds of the industry shock is credited to D.
- **Firm + industry×year:** A against its *industry peer*: $6 - 4 = 2$. The industry shock is a tide that lifts every boat in the industry; firm effects measure each boat against its own waterline.
- **Treat all of industry 1**, and firm + industry×year has nothing left to compare: D is collinear with the fixed effects. An industry-wide treatment looks exactly like an industry-wide shock.

Three effects, one story: acquisitions and ROA

Do acquisitions move profitability?

$$\text{ROA}_{ict} = \delta\,\text{Acq}_{ict} + \alpha_i + \gamma_c + \lambda_t + \varepsilon_{ict},\ \text{firm } i,\ \text{CEO } c,\ \text{year } t.$$

- **Three filters.** α_i: Apple out-earns Ford every year. γ_c: some CEOs are permanently better at running assets. λ_t: a recession pulls everyone down.
- **The mover.** Jane leaves profitable Firm A for struggling Firm B and ramps up acquisitions. Subtract B's usual ROA, Jane's usual ROA, and the year. *Did B's ROA move beyond all three when the budget spiked?* That is δ.
- **What the move buys.** Seeing Jane at A *and* at B separates γ_{Jane} from α_A.

The habit: how much of D_{it} survives?

Every effect you add absorbs part of the confounder — **and part of** D. If D barely moves within the fixed effects, your estimate rests on a handful of switchers.

```
* One command does it: sumhdfe (deHaan), a reghdfe post-estimation tool
net install sumhdfe, from("https://raw.githubusercontent.com/ed-dehaan/sumhdfe/master/src/")
reghdfe y D x, absorb(firm ind#year) vce(cluster firm)
sumhdfe                          // or standalone: sumhdfe y D x, absorb(firm ind#year)

* Panel C: groups in which D never varies -- the non-switchers, and how many obs they hold
* Panel D: pooled s.d. of D vs. within-FE s.d., and the R2 of D on each FE and on all of them

* By hand, if you want the number itself
reghdfe D, absorb(firm ind#year) residuals(D_tilde)
summarize D D_tilde              // (sd D_tilde / sd D)^2 = share of variance surviving
```

- In R: `sd(resid(feols(D ~ 1 | firm + ind^year, d))) / sd(d$D)`, squared.
- Report it next to the coefficient: *"7% of the variance in D survives the fixed effects, from 36 switching firms."* Not the same as `reghdfe`'s within R^2, which is about $\tilde{Y}$, not D.

The habit, in output

Simulated (`06a_panel_habit_sim.do`): 500 firms in 20 industries, 2011–2020; true effect of D is 0.5. A quarter of firms are always treated, 36 switch on mid-sample, the rest never.

`reghdfe`

	Coef. (s.e.)
D	0.375 (0.116)
x	0.301 (0.015)
Within R^2	0.087

`sumhdfe`, Panels C and D

	Firms where	Std. dev.			R^2 of the FE	
	it never varies	Pooled	Within	Ratio	Firm	All
y	0	1.785	0.969	54%	0.473	0.705
D	**464**	0.465	0.120	**26%**	**0.928**	0.934
x	0	1.132	0.939	83%	0.281	0.312

- **Panel C:** D never varies within 464 of 500 firms (4,640 obs.). Only the 36 switchers identify $\hat{\delta}$.
- **Panel D:** firm effects alone explain 93% of D. A quarter of its s.d. survives, so $0.26^2 \approx 7\%$ of its variance.
- The FE starve the **treatment**, not the outcome: y and x keep far more variation. $\hat{\delta} = 0.37$ (s.e. 0.12; truth 0.5) rests on 36 firms.

Three questions to ask of any panel FE result

1. **Which comparison is this?** Say it in words: *"firms that changed D in year t, against themselves before and after, net of what happened to everyone that year."* If you cannot say it, you do not know what you estimated.
2. **Is the confounder you fear actually fixed?** Write it down and ask whether it could plausibly move over your sample window. Governance quality does not sit still for fifteen years.
3. **Could the outcome be causing the treatment?** If yes, stop. You need a source of exogenous variation (Lecture 5), not more fixed effects.

- Fixed effects is not a robustness check you add at the end. It is a research design, and it makes a specific, falsifiable claim about what the confounder looks like.

Concluding Remarks

What fixed effects buys

- **Selection on unobservables, without a shock.** Matching needs you to *measure* the confounder; IV needs the world to have *moved* the treatment. Fixed effects needs only that the confounder **sit still**. That is a far weaker thing to ask of the world.
- **A claim you can state in one sentence, and partly check.** The confounder does not change — and the gap between POLS and FE tells you which way it pushed.
- **It changes the question** from *"what should I control for?"* to *"what comparison does my data allow?"* — the same unit before and after, net of what happened to everyone.
- **In finance it is the workhorse.** Firm and year effects are the default; `reghdfe` made firm + industry × year and bank + county × quarter routine. Most of what you will read, and write, is built on it.

What it costs

- **Everything fixed is gone** — including, often, the thing you cared about: founder-led, industry, state law. The design cannot say a word about them.
- **Measurement error gets worse.** Demeaning removes signal that persists and keeps noise that does not. The bias toward zero can swamp the bias you removed.
- **Strict exogeneity means no feedback.** This year's shock may not move next year's treatment. Firms exist to respond to shocks; in finance, feedback is the point, not the exception.
- **Only switchers count.** The estimate rests on the units that changed, and you did not choose them. Report how many there are.
- **Richer effects, thinner variation.** Every added effect absorbs more of D along with more of the confounder. At the limit — firm $\times$ year — nothing is left to estimate.

What fixed effects is, and is not

It removes	It does nothing about
time-invariant confounders, named or not every other time-invariant variable, wanted or not	time-varying confounders reverse causality feedback from shocks to future treatment

- One word carries the design: the confounder **does not change**. Everything strong and everything fragile about fixed effects follows from it.
- It is not a substitute for a DAG. It is what you do *after* the DAG tells you the confounder is fixed.
- Many problems in the social sciences really do have this shape. Many finance problems do not.

- The question to carry out, as always: **which comparison is this?** If you can say it in one sentence, you understand your estimate.
- **Next:** dynamic panels — when yesterday's Y is today's regressor. Then difference-in-differences, when the switching has a *date*.

References

- Abowd, J., F. Kramarz and D. Margolis (1999). High Wage Workers and High Wage Firms. *Econometrica* 67(2), 251–333.
- Andrews, M., L. Gill, T. Schank and R. Upward (2008). High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias? *Journal of the Royal Statistical Society A* 171(3), 673–697.
- Becker, G. (1968). Crime and Punishment: An Economic Approach. *JPE* 76, 169–217.
- Bell, R. and D. McCaffrey (2002). Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples. *Survey Methodology* 28(2), 169–181.
- Bertrand, M. and A. Schoar (2003). Managing with Style: The Effect of Managers on Firm Policies. *QJE* 118, 1169–1208.
- Cameron, A. C., J. Gelbach and D. Miller (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics* 90(3), 414–427.
- Cornwell, C. and P. Rupert (1997). Unobservable Individual Effects, Marriage and the Earnings of Young Men. *Economic Inquiry* 35, 285–294.
- Cornwell, C. and W. Trumbull (1994). Estimating the Economic Model of Crime with Panel Data. *Review of Economics and Statistics* 76, 360–366.
- Cunningham, S. *Causal Inference: The Remix*, §8 — this deck follows it closely.
- deHaan, E. (2021). Using and Interpreting Fixed Effects Models. Working paper, University of Washington. (Companion command: `sumhdfe`.)
- Fee, C. E., C. Hadlock and J. Pierce (2013). Managers with and without Style: Evidence Using Exogenous Variation. *RFS* 26(3), 567–601.
- Gormley, T. and D. Matsa (2014). Common Errors: How to (and Not to) Control for Unobserved Heterogeneity. *RFS* 27, 617–661.
- Greene, W. (2004). The Behaviour of the ML Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects. *Econometrics Journal* 7, 98–119.
- Hausman, J. and W. Taylor (1981). Panel Data and Unobservable Individual Effects. *Econometrica* 49, 1377–1398.
- Khwaja, A. I. and A. Mian (2008). Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market. *AER* 98(4), 1413–1442.
- MacKinnon, J. and M. Webb (2017). Wild Bootstrap Inference for Wildly Different Cluster Sizes. *Journal of Applied Econometrics* 32(2), 233–254.
- McKinnish, T. (2008). Panel Data Models and Transitory Fluctuations in the Explanatory Variable. In *Modelling and Evaluating Treatment Effects in Econometrics*.
- Roodman, D., M. Ø. Nielsen, J. MacKinnon and M. Webb (2019). Fast and Wild: Bootstrap Inference in Stata Using `boottest`. *Stata Journal* 19(1), 4–60.
- Webb, M. (2023). Reworking Wild Bootstrap-Based Inference for Clustered Errors. *Canadian Journal of Economics* 56(3), 839–858. (The six-point weights.)
- Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press. — the standard reference, including random effects, which this deck omits.