Lab 2: panel methods, tests and IV

Fixed or random effects, the tests a report needs, and an instrument put to the test

Magnus Lodefalk
Örebro University

Economics Research and Communication (NA310G)

26 October 2026



Today's goals

By the end of today you can:

1. Estimate fixed and random effects, and choose between them with a test you can defend
2. Run and read the tests your report needs: heteroskedasticity, multicollinearity, joint significance
3. Try an instrument on our data, and judge it as a referee would

From last time

Answer on your own first, then compare with a neighbour

1. IV = reduced form divided by what?
2. Lab 1's exit question: with firm fixed effects, will the AI coefficient grow or shrink?

Lab 1 in one slide: from raw data to a first regression

1. `master.do` sets `$root` and runs every do-file in order; `data/raw/` is never touched
2. `1_load_check`: one row per firm and year (`isid`), and the garbage found: zero employees, missing accounts, costs with a minus sign
3. `2_sample_vars`: industry and counties by rule, flip the sign then log, merge the occupation measure, drop step by step with each step counted
4. `3_descriptives`: the descriptive table, straight to Word
5. `4_regressions`: pooled OLS with year effects, clustered by firm, straight to Word

---

Anyone can rerun it from the raw data: that is the standard for your report.

The merge, line by line

```
rename daioe_pctl daioe_avg   // ours: firm AVERAGE
gen ssyk2012_4 = ssyk_main    // key: same name in both
merge m:1 ssyk2012_4 using ///
  "$root/data/raw/na310g_occupations_ssyk2012.dta", ///
  keepusing(daioe_pctl) keep(master match)
rename daioe_pctl daioe_main  // the MAIN occupation's
```

- Both files have a `daioe_pctl`: merge keeps ours and brings in nothing, so rename ours first
- `keepusing()`: take only this variable. `_merge == 1`: no ads that year, kept

`daioe_main` returns today, as the instrument for AI adoption (block G).

The first regression, read aloud

- AI coefficient 0.301 in a log outcome: AI firms have about 35.1 per cent higher financial expenses, same size, capital and year
- Clustered by firm, because each firm appears in up to seven years: standard error 0.157, $p = 0.055$
- An association, not an effect: firms that adopt AI differ in ways we do not see

Today: compare each firm with itself (fixed effects). Does the coefficient survive?

Is your Lab 1 complete?

1. Double-click `master.do`, then run `do "$root/check_lab1.do"`
2. All OK: you are ready for today
3. Not OK: compare your do-files with those in `lab1-catchup`, or copy its `3_descriptives.do` and `4_regressions.do` into `dofiles/` and run `master.do` again

---

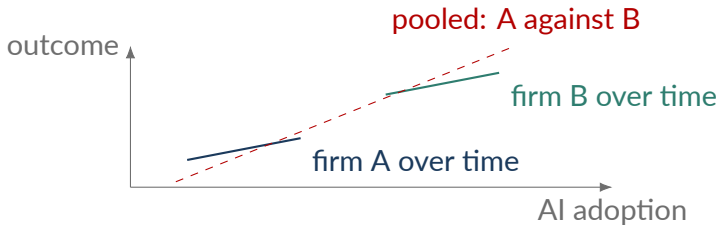
Today extends your `4_regressions.do`: it has to run first.

How today works, and words for today

- Same project: extend `4_regressions.do`, fill `5_tests_iv.do`. Opened Stata? Double-click `master.do`
- After each block: `do "$root/check_lab2.do"`. Lost? The file it names puts you back in step
- **Fixed effects:** each firm compared with itself. **Random effects:** within and between firms
- **`estimates store`:** save a result under a name, to compare it later
- **Run whole blocks:** a command split with `///` fails if you run one line of it

By 15:00 your report has its main table: pooled, fixed effects, random effects.

Fixed effects: each firm compared with itself



- Pooled OLS compares firm A with firm B: B is larger, more AI, more of everything
- Fixed effects use only the change within each firm over time; with `i.year`, two-way fixed effects

In symbols: key

$y_{it} = a_i + \lambda_t + \beta x_{it} + u_{it} \;\Rightarrow\; y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + \ldots$: a_i is gone

Stop: the AI coefficient with fixed effects

```
use "$root/data/processed/sample.dta", clear
xtset firmID year
xtreg ln_fin ai_firm ln_emp ln_cap i.year, ///
    fe vce(cluster firmID)
```

```
     ai_firm |  -.0517842   .0560016    -0.92   0.355
```

- Pooled 0.301, fixed effects −0.052: the exit question's answer is smaller; the pooled gap was a difference between firms
- Only the 285 firms (of 1,417) that change AI status identify this number
- FE remove permanent differences, not changes inside a firm (a new strategy, a boom); we still cluster

Stop: three R-squared, and which to report

- **Within:** how much of the change over time inside firms the model explains (0.07). Report this one for fixed effects
- **Between:** how well it explains differences in firms' averages
- **Overall:** all variation; not what fixed effects are built for

---

Never choose a model by its R-squared: they measure different things.

Random effects: within and between

```
xtreg ln_fin ai_firm ln_emp ln_cap i.year, ///
    re vce(cluster firmID)
```

```
     ai_firm |  -.0162605   .0555056    -0.29   0.770
```

- Uses both kinds of variation: more precise in principle (here hardly), if the firm effects a_i are unrelated to the regressors
- Between pooled (0.301) and fixed effects (-0.052): -0.016

Hausman test: which model will the data favour?

Your bet first: hands up. The answer is on the next slide.

A. Fixed effects
B. Random effects
C. The test cannot tell

Stop: the Hausman test

```
quietly xtreg ln_fin ai_firm ln_emp ln_cap i.year, fe
estimates store fe_h
quietly xtreg ln_fin ai_firm ln_emp ln_cap i.year, re
estimates store re_h
hausman fe_h re_h, sigmamore
```

- **H_0:** the firm effects are unrelated to the regressors (then RE is fine and more precise)
- p = 0.0000: reject, so use fixed effects. (`sigmamore`: both models use the same estimate of the error variance, so the test statistic cannot turn negative)

Hausman with clustered errors

```
bysort firmID: egen m_ai_firm = mean(ai_firm)
* the same for ln_emp, ln_cap and each year
xtreg ln_fin ai_firm ln_emp ln_cap m_ai_firm m_ln_emp ///
    m_ln_cap m_y2019-m_y2024 i.year, re vce(cluster firmID)
test m_ai_firm m_ln_emp m_ln_cap
```

- **Why:** `hausman` assumes ordinary standard errors; ours are clustered
- **Idea:** if a firm's *average* AI use predicts its outcome, the firm effect is tied to the regressors: RE is wrong
- **Result:** χ^2 = 120.5, reject, so fixed effects again. AI average: 0.529 (p = 0.031)

In symbols

$y_{it} = \beta x_{it} + \gamma \bar{x}_i + \ldots$, $\bar{x}_i$ = firm i's average. H_0: $\gamma = 0$ (RE is fine)

Your main table: three models side by side

	Pooled OLS	Fixed effects	Random effects
AI firm	0.301	−0.052	−0.016
	(0.157)	(0.056)	(0.056)
p-value	0.055	0.355	0.770

The `collect` block at the end of `4_regressions.do` writes this to Word (paste it from the catch-up file). In the text: which model, and the test that says so.

Three tests, and what to do with each

```
regress ln_fin ai_firm ln_emp ln_cap i.year
estat hettest, rhs iid   // rhs: each x; iid: drops normality
estat vif
```

- **Heteroskedasticity** (Breusch–Pagan): χ^2 = 22, rejects. With almost 9,000 rows almost any test does; we cluster anyway, which also allows for serial correlation
- **Multicollinearity:** the VIF of `ai_firm` is 1.06; ”above 10” is only a rule of thumb
- **Joint significance:** after the FE model, `test ln_emp ln_cap`

IV on our data: first stage, reduced form, 2SLS

```
keep if !missing(daioe_main)
regress ai_firm daioe_main ln_emp ln_cap i.year, vce(cl firmID)
regress ln_fin daioe_main ln_emp ln_cap i.year, vce(cl firmID)
ivregress 2sls ln_fin (ai_firm = daioe_main) ///
    ln_emp ln_cap i.year, vce(cluster firmID)
```

- We run it to see what a doubtful instrument looks like. `daioe_main`: the main occupation's AI exposure, a percentile rank from 0 to 100 (DAIOE)
- First stage 0.002 (F = 58.4); reduced form 0.007 (p = 0.000); 2SLS 3.694, 95% interval 1.6 to 5.8

Stop: judge it as a referee

- **Sample:** only firm-years with ads (6,330); OLS on the same sample: 0.308
- **Strength:** F = 58.4, and the reduced form (p = 0.000) is clear. But 2SLS 3.694 against OLS 0.308: about 12 times larger, an implausible size
- **Fixed effects:** Hausman said FE. With FE (`xtivreg`) the first stage vanishes (0.000, p = 0.381): exposure varies between firms, hardly within
- **Exclusion:** firms hiring for exposed occupations may also finance themselves differently (leasing, debt): exposure can move financial expenses directly, and IV then inflates

In your report: an IV only with an argument for both conditions, or a clear explanation of why none is credible. Both can earn a high grade.

The AI task: check what the AI wrote

- Open `ai_task_lab2.pdf`; answers in Word or on paper
- A chatbot wrote the paragraph on our FE, RE and Hausman results: at least four mistakes
- Find them without a chatbot; 15 minutes, then we compare

The most common report mistake is a test read backwards. Read the null hypothesis first.

Now your own question

- In your own copy (`na310g-` plus your code): pooled, FE and RE for your outcome
- Hausman and the robust version, and the three tests
- An IV only if you can argue both conditions

Lab 3 (Wed 28 Oct) and the help desks: the rest of the report. Deadline Sun 8 Nov, 16:00.

Summary

1. Fixed effects compare each firm with itself; here the pooled AI coefficient disappears, and Hausman and its robust version both say: fixed effects
2. Tests: report each, with its number and what you did; never choose a model by R-squared
3. An instrument must be strong and excluded; ours is relevant between firms only, gives an implausibly large estimate, and is doubtful on exclusion

Test yourself before next time

Without AI

1. What does a firm fixed effect remove, and what does it not remove?
2. Hausman rejects at 1 per cent. Which model do you use, and why?
3. Which R-squared do you report for fixed effects?
4. Why did our instrument lose its first stage with fixed effects?