

Örebro University · Örebro University School of Business · Economics, Economics
Research and Communication (NA310G) · Autumn 2026

AI Adoption and Financial Expenses in Large Swedish Firms

Code: EXAMPLE

About this example. It answers the lab demo question, which no student has: an outcome chosen because it shows what the methods do, not because it says anything about AI. Every number comes from the do-files in the appendix, run on the course data. The teal boxes say why each part is there and where you must choose for yourself; they are not part of a report and do not count. Text taken from this example is treated like any other copied text.

1 Introduction and research question

Firms that hire for AI skills differ from other firms in many ways, not least in size. This report asks whether firms that post job ads asking for AI skills also have higher financial expenses (interest and similar costs) than comparable firms, and whether that difference survives once each firm is compared with itself over time. The question matters for reading any firm-level comparison of AI adopters: a gap between adopters and other firms can reflect who adopts rather than what adoption does (Acemoglu et al., 2022; Engberg et al., 2024).

The purpose is to estimate the association between AI adoption and financial expenses in a panel of 1,417 large Swedish firms observed 2018–2024, and to test whether pooled OLS, fixed effects or random effects is the appropriate model. Pooled OLS shows that AI firms have about 35.1 per cent higher financial expenses than firms of the same size and capital, a difference significant at the 10 but not the 5 per cent level once standard errors are clustered by firm. With firm fixed effects the difference disappears (-0.052 , $p = 0.355$), and the Hausman test favours fixed effects. Section 2 summarises the article, section 3 states the hypotheses, sections 4 and 5 present the data and the method, sections 6 and 7 the results and the tests, and section 8 concludes.

Why it is here. The first paragraph gives the question and why it matters; the second gives the answer. A reader should know the main result after the introduction. *Your choice:* the reason your question matters depends on your outcome; do not borrow this one.

2 Summary of previous studies

Brynjolfsson and Hitt (2003) ask whether computers raise firms' productivity, a question on which industry-level studies had been inconclusive. They argue that computers pay off mainly together with slower changes in how a firm is organised, so the gains should be larger over long horizons than over short ones. With a panel of 527 large US firms observed 1987–1994, which links a survey-based measure of each firm's computer capital to its accounts, they regress the growth of productivity on the growth of computer capital, taking differences over one to seven years, with year and industry controls. Differencing compares each firm with itself, so permanent differences between firms drop out. Over one year, computers add to output about what they cost; over five to seven years, their contribution is up to five times larger. Instrumental-variables estimates and a second measure of computer capital give the same pattern, which the authors read as evidence of complementary investments in organisation rather than of measurement error.

What this section must do (about one page in your report). The paragraph above is a model of the style, not of the length: yours covers the article's question, why it matters, its contribution, theory, data, empirical strategy, results and conclusions, in your own words, for a fellow student. This example summarises an article that is not on the reading list, on purpose: yours summarises one of the starred articles, from the article itself. Note how the paragraph names the method in one sentence (differences remove permanent differences between firms, as our fixed effects do) and gives the result with its size.

3 Theory and hypotheses

Adopting AI requires investment in software, data and skills, and part of it may be financed with debt or leasing, which raises interest and similar costs. Firms that adopt are also larger and more complex, with more financial activity of every kind. Both reasons predict a positive gap between AI firms and other firms, but only the first is a consequence of adoption; the second is a difference between the firms that would show up whether or not they adopted AI. The hypotheses concern the coefficient β on the AI indicator in the model of section 5:

- $H_0: \beta = 0$, AI firms have the same financial expenses as comparable firms;
- $H_1: \beta \neq 0$, they differ (two-sided, because the theory gives reasons in both directions once permanent differences between firms are removed).

What this section must do. A mechanism from theory or an article, with references, ending in a null and an alternative about your coefficient. *Your choice:* one- or two-sided, and say why.

4 Data

The data are the course's synthetic twin of a research panel that links Serrano firm accounts to job advertisements on Platsbanken and to the occupational AI exposure of the advertised occupations (DAIOE). The twin has the variables, names, units and structure of the research panel, but the firms are simulated, so the numbers describe the twin, not Swedish firms. It covers 1,534 firms with at least 250 employees in some year, observed 2018–2024: a firm-year panel, which suits a comparison of each firm with itself over time.

The outcome is financial expenses, stored as a negative cost in the accounts; I change its sign and take the natural logarithm. Of the firm-years in the counties studied, 345 have zero financial expenses and 95 have none recorded, and the logarithm is not defined for them; I drop them, which keeps firms that borrow at all. The treatment ai_firm is 1 if the firm posted at least one job ad asking for AI skills that year. A zero mixes two kinds of firm-years: those that advertised, but not for AI, and those that did not advertise at all (32 per cent of the zeros in the sample). The required controls are the logarithms of employees and of capital (buildings and land plus machinery and equipment, summed first). Table 1 shows the sample, step by step; Appendix Table A1 defines the variables.

Table 1. From the course data to the estimation sample

Step	Dropped	Firm-years left
Course data, 2018-2024		10,492
Drop Kalmar and Västerbotten counties	217	10,275
Drop: financial expenses zero, negative or missing	440	9,835
Drop: no employees	268	9,567
Drop: no capital (buildings, land, machinery)	745	8,822

Note: all industries, all of Sweden except Kalmar and Västerbotten counties. Each step drops firm-years for which a variable in the regression is not defined.

Table 2. Descriptive statistics, estimation sample

Variable	N	Mean	Median	SD	Min	Max
Log financial expenses	8,822	7.45	7.54	2.87	0.00	16.87
AI firm (0/1)	8,822	0.11	0.00	0.32	0.00	1.00
Log employees	8,822	6.13	6.08	0.96	1.95	10.34

Variable	N	Mean	Median	SD	Min	Max
Log capital	8,822	10.30	10.55	2.89	0.69	19.84

Note: 8,822 firm-years, 1,417 firms, 2018–2024. Logarithms of amounts in thousand SEK. The median firm-year has financial expenses of 1,890 thousand SEK.

AI firm-years make up 11.3 per cent of the sample, rising from 7.8 per cent in 2018 to 12.9 per cent in 2024. Log financial expenses vary widely (standard deviation 2.87), more than size does. The correlations in Table 3 are modest: financial expenses correlate 0.10 with the AI indicator, 0.33 with employees and 0.55 with capital, and AI firms are larger (correlation 0.24 with employees). That is a first sign that the raw gap between AI firms and others partly reflects size.

Table 3. Correlations, estimation sample

	(1)	(2)	(3)	(4)
(1) Log financial expenses	1			
(2) AI firm (0/1)	0.10	1		
(3) Log employees	0.33	0.24	1	
(4) Log capital	0.55	0.07	0.37	1

Note: Pearson correlations, 8,822 firm-years.

Why it is here. The sample table lets a reader follow you from the file you got to the sample you use, and see who drops out. *Your choice:* what to do with zeros and negatives in your outcome; report how many, what you did and why. The descriptive table needs the median: the brief asks for it.

5 The econometric approach

The model is

$$\ln y_{it} = \beta \text{AI}_{it} + \gamma_1 \ln \text{employees}_{it} + \gamma_2 \ln \text{capital}_{it} + \lambda_t + a_i + u_{it},$$

where y_{it} is the financial expenses of firm i in year t , AI_{it} the AI indicator, λ_t year effects, a_i a permanent firm effect and u_{it} the error. Pooled OLS leaves a_i in the error and is consistent only if it is unrelated to AI adoption. Fixed effects remove a_i by comparing each firm with its own average, so β is identified from the 285 firms that change AI status; random effects keep the variation between firms but assume a_i is unrelated to the regressors. The Hausman test, and its version with the firms' averages that allows clustered errors (Wooldridge, 2019), tells the two apart. A firm's errors are correlated over time, so all standard errors are clustered by firm.

An instrument would have to move AI adoption (relevance) and affect financial expenses only through adoption (exclusion). A candidate is the AI exposure of the occupation the firm advertises for most (DAIOE percentile). It exists only for firm-years with ads (6,330), a selected sample. It is relevant: the first stage gives 0.0018 per percentile point, $F = 58.4$. The exclusion is not credible. Firms hiring for exposed occupations differ in how they are financed, through intangible assets, leasing and debt, so exposure can affect financial expenses directly. Two results confirm the doubt: 2SLS gives 3.69 (95 per cent interval 1.63 to 5.76), about twelve times the OLS estimate on the same sample (0.308), an implausible size; and within firms the first stage vanishes (0.0002, $p = 0.381$), because a firm's main occupation rarely changes. I therefore do not use IV.

What this section must do. Write out your equation, say what each model assumes, and why you cluster. The IV part is where a VG is earned: an instrument argued on both conditions, or a reasoned case that none is credible, as here. *Your choice:* the argument depends on your outcome; exposure may be more or less plausibly excluded for wages or sales, and you must say why.

6 Results

Table 4. AI adoption and log financial expenses

	(1) Pooled OLS	(2) Fixed effects	(3) Random effects
AI firm (0/1)	0.301* (0.157)	-0.052 (0.056)	-0.016 (0.056)
Log employees	0.421*** (0.064)	0.346*** (0.033)	0.326*** (0.030)
Log capital	0.491*** (0.021)	0.186*** (0.026)	0.363*** (0.017)
Year effects	Yes	Yes	Yes
Firm fixed effects	No	Yes	No
Firm-years	8,822	8,822	8,822
Firms	1,417	1,417	1,417
R-squared	0.32 (overall)	0.07 (within)	0.32 (overall)

Note: coefficients with standard errors clustered by firm in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All industries, Sweden except Kalmar and Västerbotten, 2018–2024.

In pooled OLS (column 1), AI firms have 0.301 log points higher financial expenses than firms with the same employment and capital in the same year: about 35.1 per cent, since $100(e^{0.301} - 1) \approx 35.1$. With standard errors clustered by firm the difference is significant at the 10 but not the 5 per cent level ($p = 0.055$); with standard errors that ignore the correlation within firms it would look highly significant ($p < 0.001$, standard error 0.085 against 0.157). A one per cent larger workforce goes with 0.421 per cent higher financial expenses, and one per cent more capital with 0.491 per cent, other things equal.

With firm fixed effects (column 2), the AI coefficient is -0.052: about -5.0 per cent, with a 95 per cent interval from -0.162 to 0.058, and not significant ($p = 0.355$). When a firm starts or stops posting AI ads, its financial expenses do not change detectably. Employment and capital keep positive and significant coefficients (0.346 and 0.186, both $p < 0.001$), and together they are significant ($F = 116.3$, $p < 0.001$). The within R-squared is 0.07: the model explains little of the change inside firms, which is common for fixed effects and no ground for choosing a model. Random effects (column 3) give -0.016 ($p = 0.770$), with an overall R-squared of 0.32.

The Hausman test has the null that the firm effects are unrelated to the regressors, so that random effects is consistent. It rejects ($\chi^2(9) = 129.4$, $p < 0.001$), and so does its version with the firms' averages and clustered errors ($\chi^2 = 120.5$, $p < 0.001$). Fixed effects is the preferred model. Controlling for firm heterogeneity matters: the pooled gap between AI firms and others reflects permanent differences between the firms, not a change that comes with adoption.

Why it is here. Every coefficient is read for sign, size with its unit, significance and the level used, and “other things equal”. Per cent for a dummy in a log outcome. The null hypothesis comes before the verdict. These are the most common reasons a report is not passed.

7 Tests for the preferred specification

Heteroskedasticity: the Breusch–Pagan test in Wooldridge's version (after pooled OLS, on the regressors) has the null of constant variance; it rejects ($\chi^2 = 21.9$, $p = 0.0093$). The standard errors in Table 4 are clustered by firm, which allows for heteroskedasticity and for correlation within firms, so no further correction is needed. Multicollinearity: the variance inflation factors of the AI indicator, employment and capital are 1.06, 1.22 and 1.15, far below the rule of thumb of 10, so collinearity

does not inflate the standard errors.

Why it is here. Each test: the null, the number, the verdict and what you did about it, in one or two sentences.

8 Concluding remarks

The report asked whether firms that advertise for AI skills have higher financial expenses than comparable firms. Pooled OLS shows a gap of about 35.1 per cent, but it is not significant at the 5 per cent level once errors are clustered by firm, and it disappears when each firm is compared with itself; the Hausman test favours that comparison. The gap is a difference between the kinds of firms that adopt AI, not a change that follows adoption.

Four limits matter. First, fixed effects remove permanent differences, not changes that coincide with adoption, such as a new strategy or rapid growth, and the estimate rests on the 285 firms that switch status. Second, `ai_firm` measures hiring for AI skills, not the use of AI, and its zeros mix firms without ads with firms that advertised for other skills. Third, causality may run the other way: a firm that borrows to expand may also start hiring for AI. Fourth, the data are a synthetic twin and the outcome was chosen to show the methods, so the results say nothing about Swedish firms. For policy, the lesson is about method rather than AI: comparisons of adopters with non-adopters, common in reports on AI and productivity, can mislead unless they account for who adopts. A study of AI and firms' finances would need real data, a measure of adoption and its timing, and a design such as an event study around the first AI ads.

Why it is here. Answer the question, state the limits honestly, and keep policy conclusions to what the results support. *Your choice:* your limits depend on your outcome and sample.

9 References

- Acemoglu, D., Autor, D., Hazell, J. and Restrepo, P. (2022). Artificial intelligence and jobs: evidence from online vacancies. *Journal of Labor Economics*, 40(S1), S293–S340.
- Brynjolfsson, E. and Hitt, L. M. (2003). Computing productivity: firm-level evidence. *Review of Economics and Statistics*, 85(4), 793–808.
- Engberg, E., Görg, H., Lodefalk, M., Javed, F., Långkvist, M., Monteiro, N. P., Kyvik Nordås, H., Pulito, G., Schroeder, S. and Tang, A. (2024). *AI Unboxed and Jobs: A Novel Measure and Firm-Level Evidence from Three Countries*. IZA Discussion Paper No. 16717. Bonn: IZA Institute of Labor Economics.
- Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1), 137–150.

Appendix (bilaga)

Table A1. Variable definitions

Variable	Definition	Source
ln_fin	Natural logarithm of financial expenses, <code>-rr09_finkostn</code> (thousand SEK)	Serrano accounts (twin)
ai_firm	1 if the firm posted at least one job ad asking for AI skills that year, else 0	Platsbanken ads, AI Monitor (twin)
ln_emp	Natural logarithm of employees, <code>bslov_antanst</code>	Serrano (twin)
ln_cap	Natural logarithm of buildings and land plus machinery and equipment, <code>br02a_byggmark + br03_maskiner</code> (thousand SEK)	Serrano (twin)
daioe_main	AI exposure (DAIOE percentile, 0–100) of the firm's most advertised occupation, merged from the occupation file	DAIOE, Engberg et al. (2024)
year	Year, 2018–2024	

Verification. I deleted `data/processed` and `output`, ran `master.do`, and got $N = 8,822$ and $b = 0.301$ as in Table 4, column 1.

AI declaration.

1. Tools: Claude (Anthropic), used by the teacher.
2. What for and how: the teacher had Claude draft this example report and its `do`-file `6_report.do` from the lab `do`-files, and checked every number against the logs. A student may not do this: the course rules allow AI to explain, debug and suggest, not to execute the analysis or write the report. No lines in the `do`-files carry `* AI:`, because the whole example is declared here.
3. Responsibility: I reviewed everything these tools produced, and I take full responsibility for the content of this report, including every line of code and every number.

Why it is here. The verification line shows the project reruns from the raw data. The declaration has three lines; yours describes your own use. Then come the `do`-files and the final log with brief comments; delete any log line that shows your name in a folder path (here the path lines are removed).

Do-files

0_settings.do

```
* =====
* PURPOSE  Settings used by every do-file
* CREATED  13 Oct 2026, EXAMPLE
* =====
* Run master.do first: it sets $root, the project folder.

cd "$root"           // work in the project folder, so every path is the same on every computer
version 18           // results stay the same in newer Stata versions (any Stata 18 or newer works)
set more off         // let output run without pausing
set seed 20261022    // anything random becomes repeatable
```

1_load_check.do

```
* =====
* PURPOSE Load the course data and check it: rows, keys, missing, zeros and negatives
* INPUTS  data/raw/na310g_data_course.dta
* OUTPUTS data/processed/checked.dta, output/logs/1_load_check.log
* CREATED 13 Oct 2026, NA310G example report (from the Lab 1 demo)
* =====
version 18
* Run master.do first: it sets $root, the project folder.
capture log close // closes a log left open if the last run stopped with an error
log using "$root/output/logs/1_load_check.log", replace text

* STOP: raw data are the evidence. We never edit them; we save copies.
* -- A1. Load the raw data
use "$root/data/raw/na310g_data_course.dta", clear

* -- A2. First look: how many rows, which variables, what do they mean?
describe, short
describe bslov_antanst rr09_finkostn br09_tillgsu ai_firm ssyk_main
codebook ai_firm, compact
* (In class also: help isid, and browse)

* -- A3. The key: one row per firm and year? If not, every later step is wrong.
isid firmID year

* -- A4. STOP: know your data. Garbage in, garbage out.
summarize bslov_antanst rr09_finkostn br09_tillgsu
count if bslov_antanst == 0 // firms with no employees: what are they?
count if missing(br09_tillgsu) // missing accounts
count if rr09_finkostn > 0 // costs are stored with a minus sign
count if rr04a_loner > 0 // salaries too: a trap for your report
tab ai_firm year if inlist(year, 2018, 2021, 2024), col

save "$root/data/processed/checked.dta", replace
log close
```

2_sample_vars.do

```
* =====
* PURPOSE Choose the sample (industry, counties) and create the variables
* INPUTS  data/processed/checked.dta, data/raw/na310g_occupations_ssyk2012.dta
* OUTPUTS data/processed/sample.dta, output/logs/2_sample_vars.log
* CREATED 13 Oct 2026, NA310G example report (from the Lab 1 demo)
* Demo outcome (nobody has it in the report): the firm's financial expenses, against
* AI adoption, size and capital. All industries, Sweden without Kalmar and Västerbotten.
* Chosen because it shows what robust and clustered standard errors and fixed effects
* do; it is a teaching example, not a finding about AI.
* =====
version 18
* Run master.do first: it sets $root, the project folder.
capture log close // closes a log left open if the last run stopped with an error
log using "$root/output/logs/2_sample_vars.log", replace text
use "$root/data/processed/checked.dta", clear

* -- B1. Industry groups from the SNI 2007 code. The data store the 2-digit division
* times 1000 (25000 = division 25). Firms outside the three groups (energy, water,
* education, health, or no code) get no group: that is by design, not a bug.
gen division = floor(bransch_sni07 / 1000)
gen industry = .
replace industry = 1 if inrange(division, 10, 33) // C manufacturing
replace industry = 2 if inrange(division, 58, 66) | inrange(division, 69, 75) // J, K, M
replace industry = 3 if inrange(division, 41, 47) | inrange(division, 49, 56) ///
| division == 68 | inrange(division, 77, 82) | inrange(division, 94, 96) // F to S
label define ind 1 "Manufacturing" 2 "Professional services" 3 "Other services"
label values industry ind
tab industry, missing

* -- B2. Geography: drop two counties (Kalmar = 8, Västerbotten = 24)
count
drop if inlist(ser_laen, 8, 24)
```

```

count

* -- B3. Variables. Costs are negative in the accounts: flip the sign, then log.
*   Logs of zero or negative values are missing: count them before they vanish.
gen finexp = -rr09_finkostn
gen ln_fin = ln(finexp)
gen ln_emp = ln(bslov_antanst)
gen capital = br02a_byggmark + br03_maskiner
gen ln_cap = ln(capital)
label variable ln_fin "Log financial expenses"
label variable ln_emp "Log employees"
label variable ln_cap "Log capital (buildings, land, machinery)"
count if missing(ln_fin)
count if missing(ln_emp)
count if missing(ln_cap)

* -- B4. Merge an occupation measure: many firm-years to one occupation (m:1).
*   The data already hold the firm's AVERAGE exposure under the same name, and merge
*   keeps the existing values: rename it first, or the merge brings in nothing.
rename daioe_pctl daioe_avg
gen ssyk2012_4 = ssyk_main // the key's name in the occupation file
merge m:1 ssyk2012_4 using "$root/data/raw/na310g_occupations_ssyk2012.dta", ///
    keepusing(daioe_pctl) keep(master match)
tab _merge // 1 = no ads that year, so no main occupation
assert _merge != 2
rename daioe_pctl daioe_main
label variable daioe_main "AI exposure (DAIOE percentile, 0-100) of the firm's most advertised occupation"
drop _merge

* -- B5. The estimation sample, one step at a time, each step counted
count
drop if missing(ln_fin)
count
drop if missing(ln_emp)
count
drop if missing(ln_cap)
count

save "$root/data/processed/sample.dta", replace
log close

```

3_descriptives.do

```

* =====
* PURPOSE Descriptive statistics and correlations for the report
* INPUTS data/processed/sample.dta
* OUTPUTS output/tables/descriptives.docx, output/logs/3_descriptives.log
* CREATED 13 Oct 2026, NA310G example report (from the Lab 1 demo)
* =====
version 18
* Run master.do first: it sets $root, the project folder.
capture log close // closes a log left open if the last run stopped with an error
log using "$root/output/logs/3_descriptives.log", replace text
use "$root/data/processed/sample.dta", clear

* -- B6. One table: N, mean, median, SD, min, max, a column each (Stata 18: dtable)
label variable ai_firm "AI firm (ad asking for AI skills)"
dtable ln_fin ln_emp ln_cap ai_firm, ///
    continuous(, statistics(count mean p50 sd min max)) nformat(%9.2f mean p50 sd min max)
collect layout (var) (result[count mean p50 sd min max])
collect export "$root/output/tables/descriptives.docx", replace

* -- B7. Correlations, with stars at the 5 per cent level
pwcorr ln_fin ln_emp ln_cap ai_firm, star(0.05)

log close

```

4_regressions.do

```

* =====
* PURPOSE Regressions for the report: pooled OLS, fixed effects, random effects
* INPUTS  data/processed/sample.dta
* OUTPUTS output/tables/regressions.docx, output/logs/4_regressions.log
* CREATED 13 Oct 2026, NA310G example report (from the Lab 1 and Lab 2 demos)
* =====
version 18
* Run master.do first: it sets $root, the project folder.
capture log close // closes a log left open if the last run stopped with an error
log using "$root/output/logs/4_regressions.log", replace text
use "$root/data/processed/sample.dta", clear
xtset firmID year

* -- C1 (Lab 1). Pooled OLS with year effects, clustered by firm
regress ln_fin ai_firm ln_emp ln_cap i.year, vce(cluster firmID)
estimates store pooled // stored: the table at the end shows all three models

* -- E1. STOP: fixed effects compare each firm with ITSELF over time. With i.year this
* is two-way FE: a common shock each year, and each firm against itself.
xtreg ln_fin ai_firm ln_emp ln_cap i.year, fe vce(cluster firmID)
estimates store fe
display _b[ai_firm]
* The three R-squared: within (over time, inside firms), between (across firms'
* averages), overall. For FE, report the within R-squared.
* We still cluster: FE remove a firm's permanent level, not the persistence of its shocks.

* -- E1b. FE learn only from firms that SWITCH AI status. How many are there?
bysort firmID: egen ever_ai = max(ai_firm)
bysort firmID: egen always_ai = min(ai_firm)
egen byte firm_tag = tag(firmID)
count if firm_tag & ever_ai == 1 & always_ai == 0 // switchers: they identify the FE estimate

* -- E2. Random effects: uses variation within AND between firms
xtreg ln_fin ai_firm ln_emp ln_cap i.year, re vce(cluster firmID)
estimates store re

* -- E3. Hausman. H0: the firm effects are unrelated to the regressors (then both are
* consistent and RE is more precise). It needs the default standard errors.
* hausman compares two stored models, so we store both
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

* -- E4. The robust version (Wooldridge 14-3, correlated random effects): add each firm's
* averages to RE, of the regressors AND of the year dummies (the panel is unbalanced).
* If the averages of the regressors matter, RE's assumption fails: use FE.
bysort firmID: egen m_ai_firm = mean(ai_firm)
bysort firmID: egen m_ln_emp = mean(ln_emp)
bysort firmID: egen m_ln_cap = mean(ln_cap)
bysort firmID: egen m_y2019 = mean(year == 2019)
bysort firmID: egen m_y2020 = mean(year == 2020)
bysort firmID: egen m_y2021 = mean(year == 2021)
bysort firmID: egen m_y2022 = mean(year == 2022)
bysort firmID: egen m_y2023 = mean(year == 2023)
bysort firmID: egen m_y2024 = mean(year == 2024)
xtreg ln_fin ai_firm ln_emp ln_cap m_ai_firm m_ln_emp m_ln_cap m_y* i.year, re vce(cluster firmID)
test m_ai_firm m_ln_emp m_ln_cap

* -- E5. One table for the report, straight to Word, ready to paste: pooled, FE, RE.
* collect is Stata's own table system (etable and dtable are built on it). Each line
* below sets one thing you can see in the table. To adapt it: change the labels and the note.
collect clear
collect create regtable
* Put each stored model into the table: its column number, its name, and two Yes/No rows
estimates restore pooled
quietly collect _r_b _r_se _r_p e(N) e(r2), tag(model[(1)] method[ols]): estimates replay
collect get firmfe = "No" yearfe = "Yes", tags(model[(1)] method[ols])
estimates restore fe
quietly collect _r_b _r_se _r_p e(N) e(r2_w) e(r2_o), tag(model[(2)] method[fe]): estimates replay
collect get firmfe = "Yes" yearfe = "Yes", tags(model[(2)] method[fe])
estimates restore re

```

```

quietly collect _r_b _r_se _r_p e(N) e(r2_w) e(r2_o), tag(model[(3)] method[re]): estimates replay
collect get firmfe = "No" yearfe = "Yes", tags(model[(3)] method[re])
* Column heads and row labels (short labels here; the variable labels stay as they are)
collect label levels method ols "Pooled OLS" fe "Fixed effects" re "Random effects", modify
collect label levels colname ai_firm "AI firm" ln_emp "Log employees" ln_cap "Log capital", modify
* R-squared: pooled OLS has one, an overall one, so it goes in the overall row;
* for FE you report the within R-squared, for RE the overall one
collect recode result r2 = r2_o
collect label levels result firmfe "Firm fixed effects" yearfe "Year effects" N "Observations" ///
    r2_w "Within R-squared" r2_o "Overall R-squared", modify
* Numbers: three decimals, standard errors in parentheses, stars on the coefficient
collect style cell result[_r_b _r_se], nformat(%9.3f)
collect style cell result[_r_se], sformat("(%s)")
collect style cell result[N], nformat(%9.0fc)
collect style cell result[r2_w r2_o], nformat(%9.2f)
collect stars _r_p 0.01 "***" 0.05 "**" 0.1 "*", attach(_r_b)
* Look: no label on coefficient and SE rows, centred numbers, no vertical line,
* a line between the coefficients and the rows below them
collect style header result[_r_b _r_se], level(hide)
collect style cell, halign(center)
collect style cell cell_type[row-header], halign(left)
collect style cell border_block, border(right, pattern(nil))
collect style cell result[firmfe], border(top, pattern(single))
* Rows: each variable (coefficient, SE below), then the other rows. Columns: (1), then the name
collect layout (colname[ai_firm ln_emp ln_cap]#result[_r_b _r_se] ///
    result[firmfe yearfe N r2_w r2_o]) (model#method)
collect title "Table 4. AI adoption and financial expenses"
collect notes "Notes: The dependent variable is log financial expenses."
collect notes "Standard errors clustered by firm in parentheses. *** p<0.01, ** p<0.05, * p<0.1."
collect preview
collect export "$root/output/tables/regressions.docx", replace

log close

```

5_tests_iv.do

```

* =====
* PURPOSE Tests on the preferred model, and an instrumental-variables attempt
* INPUTS data/processed/sample.dta
* OUTPUTS output/logs/5_tests_iv.log
* CREATED 13 Oct 2026, NA310G example report (from the Lab 2 demo)
* =====
version 18
* Run master.do first: it sets $root, the project folder.
capture log close // closes a log left open if the last run stopped with an error
log using "$root/output/logs/5_tests_iv.log", replace text
use "$root/data/processed/sample.dta", clear
xtset firmID year

* -- F1. Heteroskedasticity (Wooldridge's LM version of Breusch-Pagan): after pooled OLS
* with default standard errors. With almost 9,000 rows almost any test rejects; we use
* clustered standard errors anyway, which also allow for serial correlation.
regress ln_fin ai_firm ln_emp ln_cap i.year
estat hettest, rhs iid

* -- F2. Multicollinearity: look at the VIF of the variable you care about (ai_firm),
* not the mean, which mostly averages the year dummies. "Above 10" is a rule of thumb.
estat vif

* -- F3. Joint significance: do employment and capital matter together?
xtreg ln_fin ai_firm ln_emp ln_cap i.year, fe vce(cluster firmID)
test ln_emp ln_cap

* -- G1. IV: the AI exposure of the firm's main occupation as an instrument for ai_firm.
* We run it to see what a doubtful instrument looks like. Only firm-years with ads have a
* main occupation: a smaller, selected sample.
keep if !missing(daioe_main)
count
* OLS on the same sample, for a like-with-like comparison
regress ln_fin ai_firm ln_emp ln_cap i.year, vce(cluster firmID)
* First stage first, then the reduced form, then 2SLS

```

```

regress ai_firm daioe_main ln_emp ln_cap i.year, vce(cluster firmID)
test daioe_main
regress ln_fin daioe_main ln_emp ln_cap i.year, vce(cluster firmID)
* (with one instrument, the reduced form's test of daioe_main is the weak-IV-robust
* test of "AI has no effect")
ivregress 2sls ln_fin (ai_firm = daioe_main) ln_emp ln_cap i.year, vce(cluster firmID)
display _b[ai_firm]
estat firststage

* -- G2. Hausman said fixed effects, so a referee asks for IV with firm fixed effects too
xtivreg ln_fin ln_emp ln_cap i.year (ai_firm = daioe_main), fe vce(cluster firmID) first
display _b[ai_firm]
* STOP: relevance is tested; exclusion is an argument. Firms that hire for exposed
* occupations may also finance themselves differently (leasing, debt), so exposure
* can move financial expenses directly.

log close

```

6_report.do

```

* =====
* PURPOSE Every number and table the example report quotes, written to files the report
* is built from, so no number is typed by hand
* INPUTS data/processed/checked.dta, data/processed/sample.dta (files 1 and 2)
* OUTPUTS output/tables/report_numbers.csv, output/tables/table_sample.md,
* output/tables/table_desc.md, table_corr.md, table_reg.md, output/logs/6_report.log
* CREATED 13 Oct 2026, NA310G example report (Autumn 2026)
* NOTE You do not need a file like this one. It is the teacher's machinery for writing every
* number of the example into the text automatically, so it uses macros and file write.
* Your tables come from dtable and collect (files 3 and 4), as in the labs.
* =====
version 18
* Run master.do first: it sets $root, the project folder.
capture log close // closes a log left open if the last run stopped with an error
log using "$root/output/logs/6_report.log", replace text

* Writes one number as key,"value" to the numbers file
capture program drop put
program define put
    args key val
    file write nb ``key',"val'" _n
end

* A coefficient with stars as in the table from 4_regressions.do: *** 1%, ** 5%, * 10%
capture program drop star
program define star, rclass
    args b p
    local s ""
    if `p' < 0.10 local s "*"
    if `p' < 0.05 local s "**"
    if `p' < 0.01 local s "***"
    return local txt = string(`b', "%9.3f") + "`s'"
end

capture file close nb
file open nb using "$root/output/tables/report_numbers.csv", write replace
file write nb "key,value" _n

* -- R1. The sample, step by step (the same steps as 2_sample_vars.do)
use "$root/data/processed/checked.dta", clear
count
local n0 = r(N)
put n-row `: di %9.0fc `n0''
egen byte t = tag(firmID)
count if t
put firms-row `: di %9.0fc r(N)'
summarize year, meanonly
local ymin = r(min)
local ymax = r(max)
put year-min `ymin'
put year-max `ymax'
drop if inlist(ser_laen, 8, 24)

```

```

count
local n1 = r(N)
put drop-counties `: di %9.0fc `n0' - `n1''
put n-counties `: di %9.0fc `n1''
gen finexp = -rr09_finkostn
count if finexp == 0
put fin-zero `: di %9.0fc r(N)'
count if finexp < 0
put fin-neg `: di %9.0fc r(N)'
count if missing(finexp)
put fin-miss `: di %9.0fc r(N)'
drop if missing(ln(finexp))
count
local n2 = r(N)
put drop-fin `: di %9.0fc `n1' - `n2''
put n-fin `: di %9.0fc `n2''
drop if missing(ln(bslov_antanst))
count
local n3 = r(N)
put drop-emp `: di %9.0fc `n2' - `n3''
put n-emp `: di %9.0fc `n3''
drop if missing(ln(br02a_byggmark + br03_maskiner))
count
local n4 = r(N)
put drop-cap `: di %9.0fc `n3' - `n4''

file open tm using "$root/output/tables/table_sample.md", write replace
file write tm "| Step | Dropped | Firm-years left |" _n " |---|---|---|:" _n
file write tm "| Course data, `ymin'-'ymax' | | `: di %9.0fc `n0'' |" _n
file write tm "| Drop Kalmar and Västerbotten counties | `: di %9.0fc `n0'-'n1'' | `: di %9.0fc `n1'' |" _n
file write tm "| Drop: financial expenses zero, negative or missing | `: di %9.0fc `n1'-'n2'' | `: di %9.0fc `n2'-'n3'' |" _n
file write tm "| Drop: no employees | `: di %9.0fc `n2'-'n3'' | `: di %9.0fc `n3'' |" _n
file write tm "| Drop: no capital (buildings, land, machinery) | `: di %9.0fc `n3'-'n4'' | `: di %9.0fc `n4'' |" _n
file close tm

* -- R2. The estimation sample: size, AI shares, descriptive statistics, correlations
use "$root/data/processed/sample.dta", clear
xtset firmID year
count
put n-sample `: di %9.0fc r(N)'
egen byte t = tag(firmID)
count if t
put firms-sample `: di %9.0fc r(N)'
summarize ai_firm, meanonly
put ai-share `: di %4.1f 100*r(mean)'
summarize ai_firm if year == 2018, meanonly
put ai-share-first `: di %4.1f 100*r(mean)'
summarize ai_firm if year == 2024, meanonly
put ai-share-last `: di %4.1f 100*r(mean)'
count if ai_firm == 0
local z = r(N)
count if ai_firm == 0 & has_ads == 0
put zero-noads-pct `: di %4.0f 100*r(N)/`z''
summarize finexp, detail
put fin-median `: di %9.0fc r(p50)'
summarize ln_fin
put sd-fin `: di %4.2f r(sd)'

label variable ln_fin "Log financial expenses"
label variable ln_emp "Log employees"
label variable ln_cap "Log capital"
label variable ai_firm "AI firm (0/1)"
file open td using "$root/output/tables/table_desc.md", write replace
file write td "| Variable | N | Mean | Median | SD | Min | Max |" _n " |---|---|---|---|---|---|:" _n
foreach v in ln_fin ai_firm ln_emp ln_cap {
    quietly summarize `v', detail
    file write td "| `: variable label `v' | `: di %9.0fc r(N)' | `: di %6.2f r(mean)' | `: di %6.2f r(p50)' | `: di %6.2f r(sd)' | `: di %6.2f r(min)' | `: di %6.2f r(max)' |" _n
}
file close td
correlate ln_fin ai_firm ln_emp ln_cap

```

```

matrix C = r(C)
file open tc using "$root/output/tables/table_corr.md", write replace
file write tc "| (1) | (2) | (3) | (4) |" _n " |---|:---|:---|:---|:---| " _n
file write tc "| (1) `: variable label ln_fin' | 1 | | |" _n
file write tc "| (2) `: variable label ai_firm' | `: di %4.2f C[2,1]' | 1 | |" _n
file write tc "| (3) `: variable label ln_emp' | `: di %4.2f C[3,1]' | `: di %4.2f C[3,2]' | 1 | |" _n
file write tc "| (4) `: variable label ln_cap' | `: di %4.2f C[4,1]' | `: di %4.2f C[4,2]' | `: di %4.2f C[4,3]'
    | 1 |" _n
file close tc
put corr-fin-ai `: di %4.2f C[2,1]'
put corr-fin-emp `: di %4.2f C[3,1]'
put corr-fin-cap `: di %4.2f C[4,1]'
put corr-ai-emp `: di %4.2f C[3,2]'
put corr-emp-cap `: di %4.2f C[4,3]'

* -- R3. The three models, clustered by firm (as in 4_regressions.do)
regress ln_fin ai_firm ln_emp ln_cap i.year, vce(cluster firmID)
put b-pooled `: di %5.3f _b[ai_firm]'
put se-pooled `: di %5.3f _se[ai_firm]'
put p-pooled `: di %5.3f 2*ttail(e(df_r), abs(_b[ai_firm]/_se[ai_firm]))'
put pct-pooled `: di %4.1f 100*(exp(_b[ai_firm]) - 1)'
put b-emp-pooled `: di %5.3f _b[ln_emp]'
put b-cap-pooled `: di %5.3f _b[ln_cap]'
put r2-pooled `: di %4.2f e(r2)'
foreach v in ai_firm ln_emp ln_cap {
    star `_b[\'v\']' `=2*ttail(e(df_r), abs(_b[\'v\']/_se[\'v\']))'
    local P_\'v\' ""r(txt)""
    local Ps_\'v\' "(" + string(_se[\'v\'], "%9.3f") + ")"
}
local P_N = e(N)
local P_r2 = e(r2)
regress ln_fin ai_firm ln_emp ln_cap i.year, vce(robust)
put se-robust `: di %5.3f _se[ai_firm]'
put p-robust `: di %5.3f 2*ttail(e(df_r), abs(_b[ai_firm]/_se[ai_firm]))'
local pr = 2*ttail(e(df_r), abs(_b[ai_firm]/_se[ai_firm]))
if `pr' < 0.001 put p-robust-txt "< 0.001"
else put p-robust-txt "`=: di %5.3f `pr'""

xtreg ln_fin ai_firm ln_emp ln_cap i.year, fe vce(cluster firmID)
put b-fe `: di %5.3f _b[ai_firm]'
put se-fe `: di %5.3f _se[ai_firm]'
put p-fe `: di %5.3f 2*ttail(e(df_r), abs(_b[ai_firm]/_se[ai_firm]))'
put pct-fe `: di %4.1f 100*(exp(_b[ai_firm]) - 1)'
put lo-fe `: di %5.3f _b[ai_firm] - invttail(e(df_r), 0.025)*_se[ai_firm]'
put hi-fe `: di %5.3f _b[ai_firm] + invttail(e(df_r), 0.025)*_se[ai_firm]'
put b-emp-fe `: di %5.3f _b[ln_emp]'
put p-emp-fe `: di %5.3f 2*ttail(e(df_r), abs(_b[ln_emp]/_se[ln_emp]))'
put b-cap-fe `: di %5.3f _b[ln_cap]'
put p-cap-fe `: di %5.3f 2*ttail(e(df_r), abs(_b[ln_cap]/_se[ln_cap]))'
put r2w-fe `: di %4.2f e(r2_w)'
put firms-fe `: di %9.0fc e(N_g)'
foreach v in ai_firm ln_emp ln_cap {
    star `_b[\'v\']' `=2*ttail(e(df_r), abs(_b[\'v\']/_se[\'v\']))'
    local F_\'v\' ""r(txt)""
    local Fs_\'v\' "(" + string(_se[\'v\'], "%9.3f") + ")"
}
local F_N = e(N)
local F_r2 = e(r2_w)
local F_g = e(N_g)
test ln_emp ln_cap
put joint-F `: di %5.1f r(F)'
put joint-p `: di %5.3f r(p)'

bysort firmID: egen ever_ai = max(ai_firm)
bysort firmID: egen always_ai = min(ai_firm)
count if t & ever_ai == 1 & always_ai == 0
put switchers `: di %9.0fc r(N)'
count if t & always_ai == 1
put always-ai `: di %9.0fc r(N)'

xtreg ln_fin ai_firm ln_emp ln_cap i.year, re vce(cluster firmID)
put b-re `: di %5.3f _b[ai_firm]'
put se-re `: di %5.3f _se[ai_firm]'

```

```

put p-re `: di %5.3f 2*normal(-abs(_b[ai_firm]/_se[ai_firm]))'
put r2o-re `: di %4.2f e(r2_o)'
foreach v in ai_firm ln_emp ln_cap {
    star `_b[\'v\']' `=2*normal(-abs(_b[\'v\']/_se[\'v\']))'
    local R_\'v\' ``r(txt)''
    local Rs_\'v\' = "(" + string(_se[\'v\'], "%9.3f") + ")"
}
local R_N = e(N)
local R_r2 = e(r2_o)

file open tr using "$root/output/tables/table_reg.md", write replace
file write tr "| (1) Pooled OLS | (2) Fixed effects | (3) Random effects |" _n " |---|:---|:---|:---|" _n
foreach v in ai_firm ln_emp ln_cap {
    file write tr "| `: variable label `v' | `P_\'v\' | `F_\'v\' | `R_\'v\' |" _n
    file write tr "| `Ps_\'v\' | `Fs_\'v\' | `Rs_\'v\' |" _n
}
file write tr "| Year effects | Yes | Yes | Yes |" _n " | Firm fixed effects | No | Yes | No |" _n
file write tr "| Firm-years | `: di %9.0fc `P_N'' | `: di %9.0fc `F_N'' | `: di %9.0fc `R_N'' |" _n
file write tr "| Firms | `: di %9.0fc `F_g'' | `: di %9.0fc `F_g'' | `: di %9.0fc `F_g'' |" _n
file write tr "| R-squared | `: di %4.2f `P_r2'' (overall) | `: di %4.2f `F_r2'' (within) | `: di %4.2f `R_r2'' (
    overall) |" _n
file close tr

* -- R4. Which model: Hausman (default standard errors) and its robust version
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
put hausman-chi2 `: di %5.1f r(chi2)'
put hausman-df `r(df)'
put hausman-p `: di %6.4f r(p)'
bysort firmID: egen m_ai_firm = mean(ai_firm)
bysort firmID: egen m_ln_emp = mean(ln_emp)
bysort firmID: egen m_ln_cap = mean(ln_cap)
forvalues y = 2019/2024 {
    bysort firmID: egen m_y`y' = mean(year == `y')
}
xtreg ln_fin ai_firm ln_emp ln_cap m_ai_firm m_ln_emp m_ln_cap m_y* i.year, re vce(cluster firmID)
test m_ai_firm m_ln_emp m_ln_cap
put cre-chi2 `: di %5.1f r(chi2)'
put cre-p `: di %6.4f r(p)'

* -- R5. Tests for the preferred specification
regress ln_fin ai_firm ln_emp ln_cap i.year
estat hettest, rhs iid
put het-chi2 `: di %5.1f r(chi2)'
put het-p `: di %6.4f r(p)'
estat vif
* The VIFs estat vif prints above, from each regressor's auxiliary regression (year dummies included)
quietly regress ai_firm ln_emp ln_cap i.year
put vif-ai `: di %4.2f 1/(1 - e(r2))'
quietly regress ln_emp ai_firm ln_cap i.year
put vif-emp `: di %4.2f 1/(1 - e(r2))'
quietly regress ln_cap ai_firm ln_emp i.year
put vif-cap `: di %4.2f 1/(1 - e(r2))'

* -- R6. The instrument considered in section 5 (as in 5_tests_iv.do)
keep if !missing(daioe_main)
count
put n-iv `: di %9.0fc r(N)'
regress ln_fin ai_firm ln_emp ln_cap i.year, vce(cluster firmID)
put b-ols-iv `: di %5.3f _b[ai_firm]'
regress ai_firm daioe_main ln_emp ln_cap i.year, vce(cluster firmID)
put fs-b `: di %6.4f _b[daioe_main]'
test daioe_main
put fs-F `: di %5.1f r(F)'
regress ln_fin daioe_main ln_emp ln_cap i.year, vce(cluster firmID)
put rf-b `: di %6.4f _b[daioe_main]'
put rf-p `: di %6.4f 2*ttail(e(df_r), abs(_b[daioe_main]/_se[daioe_main]))'
ivregress 2sls ln_fin (ai_firm = daioe_main) ln_emp ln_cap i.year, vce(cluster firmID)
put b-iv `: di %5.2f _b[ai_firm]'
put se-iv `: di %5.2f _se[ai_firm]'

```

```

put lo-iv `: di %5.2f _b[ai_firm] - invnormal(0.975)*_se[ai_firm]'
put hi-iv `: di %5.2f _b[ai_firm] + invnormal(0.975)*_se[ai_firm]'
xtreg ai_firm daioe_main ln_emp ln_cap i.year, fe vce(cluster firmID)
put fefs-b `: di %6.4f _b[daioe_main]'
put fefs-p `: di %5.3f 2*ttail(e(df_r), abs(_b[daioe_main]/_se[daioe_main]))'

file close nb
log close

```

Final log

1_load_check.log

```

-----
name: <unnamed>
log type: text
opened on: 16 Oct 2026, 17:05:57

.
. * STOP: raw data are the evidence. We never edit them; we save copies.
. * -- A1. Load the raw data
. use "$root/data/raw/na310g_data_course.dta", clear
(SYNTHETIC NA310G firm panel 2018-2024, dataset course - not real firms)

.
. * -- A2. First look: how many rows, which variables, what do they mean?
. describe, short

Contains data from .../project/data/raw/na310g_data_course.dta
Observations:      10,492      SYNTHETIC NA310G firm panel
                                2018-2024, dataset course - not
                                real firms
Variables:          56      13 Oct 2026 14:52
Sorted by:

. describe bslov_antanst rr09_finkostn br09_tillgsu ai_firm ssyk_main

Variable      Storage   Display   Value
name          type      format    label      Variable label
-----
bslov_antanst  long      %12.0g    Employees
rr09_finkostn  double   %10.0g    Financial expenses
br09_tillgsu   double   %10.0g    Total assets
ai_firm        byte      %8.0g     AI adoption proxy: 1 if at least
                                one ad mentions AI skills
ssyk_main      double   %10.0g    Most advertised occupation this
                                year (SSYK 2012, 4-digit)

. codebook ai_firm, compact

Variable      Obs Unique      Mean  Min  Max  Label
-----
ai_firm      10492      2  .1097026  0   1  AI adoption proxy: 1 if at leas...

. * (In class also: help isid, and browse)
.
. * -- A3. The key: one row per firm and year? If not, every later step is wron
> g.
. isid firmID year

.
. * -- A4. STOP: know your data. Garbage in, garbage out.
. summarize bslov_antanst rr09_finkostn br09_tillgsu

Variable |      Obs      Mean  Std. dev.      Min      Max
-----+-----
bslov_anta~t |    10,492    743.923    1152.746         0    31026
rr09_finko~n |    10,396   -59996.06    488980.4   -2.13e+07         0
br09_tillgsu |    10,396    4088194    1.74e+07         48    4.93e+08

```

```

. count if bslov_antanst == 0          // firms with no employees: what are th
> ey?
390

. count if missing(br09_tillgsu)      // missing accounts
96

. count if rr09_finkostn > 0          // costs are stored with a minus sign
96

. count if rr04a_loner > 0            // salaries too: a trap for your repor
> t
129

. tab ai_firm year if inlist(year, 2018, 2021, 2024), col

+-----+
| Key          |
+-----+
| frequency    |
| column percentage |
+-----+

      AI |
adoption |
proxy: 1 |
if at    |
least one |
ad       |
mentions |
AI skills |
Year      |
2018      |
2021      |
2024      |
Total     |
+-----+
0 |      1,377      1,337      1,300 |      4,014
  |      92.05      88.78      87.31 |      89.38
+-----+
1 |      119        169        189 |      477
  |       7.95      11.22      12.69 |     10.62
+-----+
Total |      1,496      1,506      1,489 |      4,491
      |     100.00     100.00     100.00 |     100.00

.
. save "$root/data/processed/checked.dta", replace
(file
.../project/data/processed/checked.dta not found)
file
.../project/data/processed/checked.dta saved

. log close
name: <unnamed>
log type: text
closed on: 16 Oct 2026, 17:05:57
-----

```

2_sample_vars.log

```

-----
name: <unnamed>
log type: text
opened on: 16 Oct 2026, 17:05:57

. use "$root/data/processed/checked.dta", clear
(SYNTHETIC NA310G firm panel 2018-2024, dataset course - not real firms)

.
. * -- B1. Industry groups from the SNI 2007 code. The data store the 2-digit d
> ivision
. * times 1000 (25000 = division 25). Firms outside the three groups (energy
> , water,
. * education, health, or no code) get no group: that is by design, not a bu
> g.

```

```

. gen division = floor(bransch_sni07 / 1000)
(68 missing values generated)

. gen industry = .
(10,492 missing values generated)

. replace industry = 1 if inrange(division, 10, 33)
> // C manufacturing
(2,133 real changes made)

. replace industry = 2 if inrange(division, 58, 66) | inrange(division, 69, 75)
> // J, K, M
(2,096 real changes made)

. replace industry = 3 if inrange(division, 41, 47) | inrange(division, 49, 56)
> ///
> | division == 68 | inrange(division, 77, 82) | inrange(
> division, 94, 96) // F to S
(4,429 real changes made)

. label define ind 1 "Manufacturing" 2 "Professional services" 3 "Other service
> s"

. label values industry ind

. tab industry, missing

```

industry	Freq.	Percent	Cum.
Manufacturing	2,133	20.33	20.33
Professional services	2,096	19.98	40.31
Other services	4,429	42.21	82.52
.	1,834	17.48	100.00
Total	10,492	100.00	

```

.
. * -- B2. Geography: drop two counties (Kalmar = 8, Västerbotten = 24)
. count
10,492

. drop if inlist(ser_laen, 8, 24)
(217 observations deleted)

. count
10,275

.
. * -- B3. Variables. Costs are negative in the accounts: flip the sign, then l
> og.
. * Logs of zero or negative values are missing: count them before they vani
> sh.
. gen finexp = -rr09_finkostn
(95 missing values generated)

. gen ln_fin = ln(finexp)
(440 missing values generated)

. gen ln_emp = ln(bslov_antanst)
(388 missing values generated)

. gen capital = br02a_byggmark + br03_maskiner
(95 missing values generated)

. gen ln_cap = ln(capital)
(909 missing values generated)

. label variable ln_fin "Log financial expenses"

. label variable ln_emp "Log employees"

. label variable ln_cap "Log capital (buildings, land, machinery)"

```

```

. count if missing(ln_fin)
440

. count if missing(ln_emp)
388

. count if missing(ln_cap)
909

.
. * -- B4. Merge an occupation measure: many firm-years to one occupation (m:1)
> .
. * The data already hold the firm's AVERAGE exposure under the same name, a
> nd merge
. * keeps the existing values: rename it first, or the merge brings in nothi
> ng.
. rename daioe_pctl daioe_avg

. gen ssyk2012_4 = ssyk_main // the key's name in the oc
> cupation file
(3,147 missing values generated)

. merge m:1 ssyk2012_4 using "$root/data/raw/na310g_occupations_ssyk2012.dta",
> ///
> keepusing(daioe_pctl) keep(master match)

Result                                Number of obs
-----
Not matched                          3,147
  from master                        3,147  (_merge==1)
  from using                          0  (_merge==2)

Matched                              7,128  (_merge==3)
-----

. tab _merge // 1 = no ads that year, so
> no main occupation

Matching result from |
merge | Freq. Percent Cum.
-----+-----
Master only (1) | 3,147 30.63 30.63
Matched (3) | 7,128 69.37 100.00
-----+-----
Total | 10,275 100.00

. assert _merge != 2

. rename daioe_pctl daioe_main

. label variable daioe_main "AI exposure (DAIOE percentile, 0-100) of the firm"
> s most advertised occupation"

. drop _merge

.
. * -- B5. The estimation sample, one step at a time, each step counted
. count
10,275

. drop if missing(ln_fin)
(440 observations deleted)

. count
9,835

. drop if missing(ln_emp)
(268 observations deleted)

. count
9,567

. drop if missing(ln_cap)

```

```

(745 observations deleted)

. count
      8,822

.
. save "$root/data/processed/sample.dta", replace
(file
  .../project/data/processed/sample.dta not found)
file
  .../project/data/processed/sample.dta saved

. log close
  name: <unnamed>
  log type: text
  closed on: 16 Oct 2026, 17:05:57

```

3_descriptives.log

```

-----
  name: <unnamed>
  log type: text
  opened on: 16 Oct 2026, 17:05:57

. use "$root/data/processed/sample.dta", clear
(SYNTHETIC NA310G firm panel 2018–2024, dataset course – not real firms)

.
. * -- B6. One table: N, mean, median, SD, min, max, a column each (Stata 18: d
> table)
. label variable ai_firm "AI firm (ad asking for AI skills)"

. dtable ln_fin ln_emp ln_cap ai_firm, ///
>   continuous(, statistics(count mean p50 sd min max)) nformat(%9.2f mean p5
> 0 sd min max)

```

Summary					
N					8,822
Log financial expenses	8,822	7.45	7.54	(2.87)	0.00 16.87
Log employees	8,822	6.13	6.08	(0.96)	1.95 10.34
Log capital (buildings, land, machinery)	8,822	10.30	10.55	(2.89)	0.69 19.84
AI firm (ad asking for AI skills)	8,822	0.11	0.00	(0.32)	0.00 1.00

```

. collect layout (var) (result[count mean p50 sd min max])

```

```

Collection: DTable
  Rows: var
  Columns: result[count mean p50 sd min max]
  Table 1: 4 x 6

```

Minimum value Maximum value		Number of nonmissing values		Mean 50th percentile	Standard deviation
Log financial expenses		8,822	7.45	7.54	(2.87)
0.00	16.87				
Log employees		8,822	6.13	6.08	(0.96)
1.95	10.34				
Log capital (buildings, land, machinery)		8,822	10.30	10.55	(2.89)
0.69	19.84				
AI firm (ad asking for AI skills)		8,822	0.11	0.00	(0.32)
0.00	1.00				

```
. collect export "$root/output/tables/descriptives.docx", replace
(collection DTable exported to file
.../project/output/tables/descriptives.docx)
```

```
. * -- B7. Correlations, with stars at the 5 per cent level
. pwcorr ln_fin ln_emp ln_cap ai_firm, star(0.05)
```

	ln_fin	ln_emp	ln_cap	ai_firm
ln_fin	1.0000			
ln_emp	0.3313*	1.0000		
ln_cap	0.5485*	0.3657*	1.0000	
ai_firm	0.1047*	0.2408*	0.0735*	1.0000

```
. log close
name: <unnamed>
log type: text
closed on: 16 Oct 2026, 17:05:59
```

4_regressions.log

```
name: <unnamed>
log type: text
opened on: 16 Oct 2026, 17:05:59
```

```
. use "$root/data/processed/sample.dta", clear
(SYNTHETIC NA310G firm panel 2018–2024, dataset course – not real firms)
```

```
. xtset firmID year
```

```
Panel variable: firmID (unbalanced)
Time variable: year, 2018 to 2024, but with gaps
Delta: 1 unit
```

```
. * -- C1 (Lab 1). Pooled OLS with year effects, clustered by firm
. regress ln_fin ai_firm ln_emp ln_cap i.year, vce(cluster firmID)
```

```
Linear regression               Number of obs   =      8,822
                               F(9, 1416)         =     108.13
                               Prob > F            =     0.0000
                               R-squared            =     0.3225
                               Root MSE         =     2.3675
```

(Std. err. adjusted for 1,417 clusters in firmID)

ln_fin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ai_firm	.3007456	.156736	1.92	0.055	–.0067141	.6082053
ln_emp	.4210508	.0643936	6.54	0.000	.2947338	.5473679
ln_cap	.4908312	.020946	23.43	0.000	.4497427	.5319197
year						
2019	.0144356	.0402789	0.36	0.720	–.0645771	.0934483
2020	.0814533	.0446086	1.83	0.068	–.0060527	.1689593
2021	.0778108	.0483458	1.61	0.108	–.0170264	.1726479
2022	.1698464	.0501316	3.39	0.001	.0715063	.2681865
2023	.1937819	.0506637	3.82	0.000	.0943978	.2931659
2024	.2298345	.051312	4.48	0.000	.1291788	.3304902
_cons	–.3289511	.3694462	–0.89	0.373	–1.053672	.3957696

```
. estimates store pooled // stored: the table at the end show
> s all three models
```

```

.
. * -- E1. STOP: fixed effects compare each firm with ITSELF over time. With i.
> year this
. * is two-way FE: a common shock each year, and each firm against itself.
. xtreg ln_fin ai_firm ln_emp ln_cap i.year, fe vce(cluster firmID)

Fixed-effects (within) regression              Number of obs   =      8,822
Group variable: firmID                       Number of groups  =      1,417

R-squared:                                  Obs per group:
    Within = 0.0657                          min =          1
    Between = 0.3434                         avg =          6.2
    Overall = 0.3047                         max =          7

corr(u_i, Xb) = 0.3535                      F(9, 1416)       =      46.19
                                              Prob > F          =      0.0000

```

(Std. err. adjusted for 1,417 clusters in firmID)

ln_fin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ai_firm	-.0517842	.0560016	-0.92	0.355	-.1616392	.0580708
ln_emp	.3462222	.032997	10.49	0.000	.2814941	.4109504
ln_cap	.1859054	.0263253	7.06	0.000	.1342646	.2375462
year						
2019	.0639915	.0361611	1.77	0.077	-.0069436	.1349267
2020	.1359775	.0385612	3.53	0.000	.0603344	.2116207
2021	.1628609	.0398493	4.09	0.000	.0846908	.2410309
2022	.2623581	.0403543	6.50	0.000	.1831975	.3415187
2023	.3237607	.0415182	7.80	0.000	.2423168	.4052046
2024	.3738758	.0420705	8.89	0.000	.2913487	.456403
_cons	3.232015	.2782346	11.62	0.000	2.686219	3.777811
sigma_u	2.4062081					
sigma_e	.95849744					
rho	.86305286	(fraction of variance due to u_i)				

```

. estimates store fe

. display _b[ai_firm]
-.05178421

. * The three R-squared: within (over time, inside firms), between (across f
> irms'
. * averages), overall. For FE, report the within R-squared.
. * We still cluster: FE remove a firm's permanent level, not the persistenc
> e of its shocks.

.
. * -- E1b. FE learn only from firms that SWITCH AI status. How many are there?
. bysort firmID: egen ever_ai = max(ai_firm)

. bysort firmID: egen always_ai = min(ai_firm)

. egen byte firm_tag = tag(firmID)

. count if firm_tag & ever_ai == 1 & always_ai == 0    // switchers: they iden
> tify the FE estimate
285

.
. * -- E2. Random effects: uses variation within AND between firms
. xtreg ln_fin ai_firm ln_emp ln_cap i.year, re vce(cluster firmID)

Random-effects GLS regression              Number of obs   =      8,822
Group variable: firmID                       Number of groups  =      1,417

R-squared:                                  Obs per group:
    Within = 0.0617                          min =          1
    Between = 0.3543                         avg =          6.2

```

max = 7

```
Wald chi2(9)      =      965.91
Prob > chi2       =      0.0000
```

	ln_fin	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
	ai_firm	-.0162605	.05555056	-0.29	0.770	-.1250495	.0925285
	ln_emp	.3260244	.0295881	11.02	0.000	.2680328	.3840159
	ln_cap	.3626923	.0171697	21.12	0.000	.3290403	.3963442
	year						
	2019	.0509667	.0361791	1.41	0.159	-.019943	.1218763
	2020	.1232491	.0387679	3.18	0.001	.0472655	.1992328
	2021	.1429643	.0400893	3.57	0.000	.0643907	.2215379
	2022	.2368771	.0404113	5.86	0.000	.1576724	.3160817
	2023	.2753829	.0411827	6.69	0.000	.1946662	.3560995
	2024	.3277374	.0421059	7.78	0.000	.2452114	.4102635
	_cons	1.532338	.221624	6.91	0.000	1.097963	1.966713
	sigma_u	2.1731576					
	sigma_e	.95849744					
	rho	.8371453	(fraction of variance due to u_i)				

```
. * -- E3. Hausman. H0: the firm effects are unrelated to the regressors (then
> both are
. *      consistent and RE is more precise). It needs the default standard errors
> .
. *      hausman compares two stored models, so we store both
. 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
```

b = Consistent under H_0 and H_a ; obtained from xtreg.
B = Inconsistent under H_a , efficient under H_0 ; obtained from xtreg.

```
. * -- E4. The robust version (Wooldridge 14-3, correlated random effects): add
```

```

> each firm's
. * averages to RE, of the regressors AND of the year dummies (the panel is
> unbalanced).
. * If the averages of the regressors matter, RE's assumption fails: use FE.
. bysort firmID: egen m_ai_firm = mean(ai_firm)

. bysort firmID: egen m_ln_emp = mean(ln_emp)

. bysort firmID: egen m_ln_cap = mean(ln_cap)

. bysort firmID: egen m_y2019 = mean(year == 2019)

. bysort firmID: egen m_y2020 = mean(year == 2020)

. bysort firmID: egen m_y2021 = mean(year == 2021)

. bysort firmID: egen m_y2022 = mean(year == 2022)

. bysort firmID: egen m_y2023 = mean(year == 2023)

. bysort firmID: egen m_y2024 = mean(year == 2024)

. xtreg ln_fin ai_firm ln_emp ln_cap m_ai_firm m_ln_emp m_ln_cap m_y* i.year, r
> e vce(cluster firmID)

```

```

Random-effects GLS regression           Number of obs   =      8,822
Group variable: firmID                  Number of groups  =      1,417

```

```

R-squared:                               Obs per group:
    Within = 0.0657                        min =          1
    Between = 0.3583                       avg =         6.2
    Overall = 0.3268                       max =          7

```

```

corr(u_i, X) = 0 (assumed)                Wald chi2(18)    =    1300.68
                                           Prob > chi2      =      0.0000

```

(Std. err. adjusted for 1,417 clusters in firmID)

	ln_fin	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ai_firm		-.0517842	.0560302	-0.92	0.355	-.1616014	.058033
ln_emp		.3462222	.0330138	10.49	0.000	.2815163	.4109281
ln_cap		.1859054	.0263387	7.06	0.000	.1342824	.2375284
m_ai_firm		.5289002	.2457536	2.15	0.031	.047232	1.010568
m_ln_emp		.1048149	.0816343	1.28	0.199	-.0551853	.2648152
m_ln_cap		.3104624	.0336084	9.24	0.000	.2445912	.3763336
m_y2019		.1188686	1.376776	0.09	0.931	-2.579563	2.8173
m_y2020		-1.167846	1.164678	-1.00	0.316	-3.450573	1.114882
m_y2021		-.8367355	1.113546	-0.75	0.452	-3.019245	1.345774
m_y2022		.8953286	1.21383	0.74	0.461	-1.483735	3.274392
m_y2023		-.9786595	1.175671	-0.83	0.405	-3.282933	1.325614
m_y2024		-.5964793	.882118	-0.68	0.499	-2.325399	1.13244
year							
2019		.0639915	.0361796	1.77	0.077	-.0069192	.1349022
2020		.1359775	.0385809	3.52	0.000	.0603604	.2115946
2021		.1628609	.0398697	4.08	0.000	.0847177	.241004
2022		.2623581	.0403749	6.50	0.000	.1832247	.3414914
2023		.3237607	.0415395	7.79	0.000	.2423448	.4051765
2024		.3738758	.042092	8.88	0.000	.2913771	.4563746
_cons		-.321528	.7777685	-0.41	0.679	-1.845926	1.20287
sigma_u		2.1731576					
sigma_e		.95849744					
rho		.8371453					(fraction of variance due to u_i)

```

. test m_ai_firm m_ln_emp m_ln_cap

```

```

( 1) m_ai_firm = 0
( 2) m_ln_emp = 0

```

```

( 3)  m_ln_cap = 0

      chi2( 3) = 120.46
      Prob > chi2 = 0.0000

.
. * -- E5. One table for the report, straight to Word, ready to paste: pooled,
> FE, RE.
. * collect is Stata's own table system (etable and dtable are built on it).
> Each line
. * below sets one thing you can see in the table. To adapt it: change the l
> abels and the note.
. collect clear

. collect create regtable
(current collection is regtable)

. * Put each stored model into the table: its column number, its name, and t
> wo Yes/No rows
. estimates restore pooled
(results pooled are active now)

. quietly collect _r_b _r_se _r_p e(N) e(r2), tag(model[(1)] method[ols]): esti
> mates replay

. collect get firmfe = "No" yearfe = "Yes", tags(model[(1)] method[ols])

. estimates restore fe
(results fe are active now)

. quietly collect _r_b _r_se _r_p e(N) e(r2_w) e(r2_o), tag(model[(2)] method[f
> e]): estimates replay

. collect get firmfe = "Yes" yearfe = "Yes", tags(model[(2)] method[fe])

. estimates restore re
(results re are active now)

. quietly collect _r_b _r_se _r_p e(N) e(r2_w) e(r2_o), tag(model[(3)] method[r
> e]): estimates replay

. collect get firmfe = "No" yearfe = "Yes", tags(model[(3)] method[re])

. * Column heads and row labels (short labels here; the variable labels stay
> as they are)
. collect label levels method ols "Pooled OLS" fe "Fixed effects" re "Random ef
> fects", modify

. collect label levels colname ai_firm "AI firm" ln_emp "Log employees" ln_cap
> "Log capital", modify

. * R-squared: pooled OLS has one, an overall one, so it goes in the overall
> row;
. * for FE you report the within R-squared, for RE the overall one
. collect recode result r2 = r2_o
(2 items recoded in collection regtable)

. collect label levels result firmfe "Firm fixed effects" yearfe "Year effects"
> N "Observations" ///
> r2_w "Within R-squared" r2_o "Overall R-squared", modify

. * Numbers: three decimals, standard errors in parentheses, stars on the co
> efficient
. collect style cell result[_r_b _r_se], nformat(%9.3f)

. collect style cell result[_r_se], sformat("(%s)")

. collect style cell result[N], nformat(%9.0fc)

. collect style cell result[r2_w r2_o], nformat(%9.2f)

. collect stars _r_p 0.01 "***" 0.05 "**" 0.1 "*", attach(_r_b)

```

```

. * Look: no label on coefficient and SE rows, centred numbers, no vertical
> line,
. * a line between the coefficients and the rows below them
. collect style header result[_r_b _r_se], level(hide)

. collect style cell, halign(center)

. collect style cell cell_type[row-header], halign(left)

. collect style cell border_block, border(right, pattern(nil))

. collect style cell result[firmfe], border(top, pattern(single))

. * Rows: each variable (coefficient, SE below), then the other rows. Column
> s: (1), then the name
. collect layout (colname[ai_firm ln_emp ln_cap]#result[_r_b _r_se] ///
> result[firmfe yearfe N r2_w r2_o]) (model#method)

```

Collection: regtable

Rows: colname[ai_firm ln_emp ln_cap]#result[_r_b _r_se] result[firmfe
yearfe N r2_w r2_o]

Columns: model#method

Table 1: 11 x 3

	(1)	(2)	(3)
	Pooled OLS	Fixed effects	Random effects
AI firm	0.301* (0.157)	-0.052 (0.056)	-0.016 (0.056)
Log employees	0.421*** (0.064)	0.346*** (0.033)	0.326*** (0.030)
Log capital	0.491*** (0.021)	0.186*** (0.026)	0.363*** (0.017)
Firm fixed effects	No	Yes	No
Year effects	Yes	Yes	Yes
Observations	8,822	8,822	8,822
Within R-squared		0.07	0.06
Overall R-squared	0.32	0.30	0.32

```

. collect title "Table 4. AI adoption and financial expenses"

. collect notes "Notes: The dependent variable is log financial expenses."

. collect notes "Standard errors clustered by firm in parentheses. *** p<0.01,
> ** p<0.05, * p<0.1."

. collect preview

```

Table 4. AI adoption and financial expenses

	(1)	(2)	(3)
	Pooled OLS	Fixed effects	Random effects
AI firm	0.301* (0.157)	-0.052 (0.056)	-0.016 (0.056)
Log employees	0.421*** (0.064)	0.346*** (0.033)	0.326*** (0.030)
Log capital	0.491*** (0.021)	0.186*** (0.026)	0.363*** (0.017)
Firm fixed effects	No	Yes	No
Year effects	Yes	Yes	Yes
Observations	8,822	8,822	8,822
Within R-squared		0.07	0.06
Overall R-squared	0.32	0.30	0.32

Notes: The dependent variable is log financial expenses.
Standard errors clustered by firm in parentheses. ***
p<0.01, ** p<0.05, * p<0.1.

```
. collect export "$root/output/tables/regressions.docx", replace
(collection regtable exported to file
.../project/output/tables/regressions.docx)
```

```
.
. log close
  name: <unnamed>
  log type: text
  closed on: 16 Oct 2026, 17:06:01
```

5_tests_iv.log

```
-----
  name: <unnamed>
  log type: text
  opened on: 16 Oct 2026, 17:06:01
```

```
. use "$root/data/processed/sample.dta", clear
(SYNTHETIC NA310G firm panel 2018-2024, dataset course - not real firms)
```

```
. xtset firmID year
```

```
Panel variable: firmID (unbalanced)
Time variable: year, 2018 to 2024, but with gaps
Delta: 1 unit
```

```
.
. * -- F1. Heteroskedasticity (Wooldridge's LM version of Breusch-Pagan): after
> pooled OLS
. * with default standard errors. With almost 9,000 rows almost any test rej
> ects; we use
. * clustered standard errors anyway, which also allow for serial correlatio
> n.
. regress ln_fin ai_firm ln_emp ln_cap i.year
```

Source	SS	df	MS	Number of obs	=	8,822
				F(9, 8812)	=	466.13
Model	23513.6589	9	2612.62876	Prob > F	=	0.0000
Residual	49390.7663	8,812	5.60494397	R-squared	=	0.3225
				Adj R-squared	=	0.3218
Total	72904.4252	8,821	8.26487078	Root MSE	=	2.3675

ln_fin	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
ai_firm	.3007456	.0822276	3.66	0.000	.1395604	.4619308
ln_emp	.4210508	.0289166	14.56	0.000	.3643675	.4777341
ln_cap	.4908312	.0093699	52.38	0.000	.4724641	.5091983
year						
2019	.0144356	.0938755	0.15	0.878	-.1695824	.1984535
2020	.0814533	.0940161	0.87	0.386	-.1028402	.2657467
2021	.0778108	.0943769	0.82	0.410	-.10719	.2628115
2022	.1698464	.0944169	1.80	0.072	-.0152328	.3549256
2023	.1937819	.0940475	2.06	0.039	.0094267	.378137
2024	.2298345	.0937793	2.45	0.014	.0460053	.4136638
_cons	-.3289511	.1769863	-1.86	0.063	-.6758856	.0179834

```
. estat hettest, rhs iid-
```

```
BreuschPagan/-CookWeisberg test for heteroskedasticity
Assumption: i.i.d. error terms
Variables: All independent variables
```

```
H0: Constant variance
```

```
chi2(9) = 21.86
Prob > chi2 = 0.0093
```

```

. * -- F2. Multicollinearity: look at the VIF of the variable you care about (a
> i_firm),
. * not the mean, which mostly averages the year dummies. "Above 10" is a ru
> le of thumb.
. estat vif

```

Variable	VIF	1/VIF
-----+-----		
ai_firm	1.06	0.939042
ln_emp	1.22	0.818220
ln_cap	1.15	0.865844
year		
2019	1.70	0.587326
2020	1.70	0.589069
2021	1.69	0.590077
2022	1.70	0.588388
2023	1.70	0.587116
2024	1.71	0.583925
-----+-----		
Mean VIF	1.52	

```

.
. * -- F3. Joint significance: do employment and capital matter together?
. xtreg ln_fin ai_firm ln_emp ln_cap i.year, fe vce(cluster firmID)

Fixed-effects (within) regression              Number of obs   =      8,822
Group variable: firmID                        Number of groups  =      1,417

R-squared:                                    Obs per group:
    Within = 0.0657                               min =          1
    Between = 0.3434                              avg =          6.2
    Overall = 0.3047                               max =          7

                                         F(9, 1416)      =      46.19
corr(u_i, Xb) = 0.3535                     Prob > F         =      0.0000

```

(Std. err. adjusted for 1,417 clusters in firmID)

	ln_fin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]
-----+-----						
ai_firm		-.0517842	.0560016	-0.92	0.355	-.1616392 .0580708
ln_emp		.3462222	.032997	10.49	0.000	.2814941 .4109504
ln_cap		.1859054	.0263253	7.06	0.000	.1342646 .2375462
year						
2019		.0639915	.0361611	1.77	0.077	-.0069436 .1349267
2020		.1359775	.0385612	3.53	0.000	.0603344 .2116207
2021		.1628609	.0398493	4.09	0.000	.0846908 .2410309
2022		.2623581	.0403543	6.50	0.000	.1831975 .3415187
2023		.3237607	.0415182	7.80	0.000	.2423168 .4052046
2024		.3738758	.0420705	8.89	0.000	.2913487 .456403
_cons		3.232015	.2782346	11.62	0.000	2.686219 3.777811
-----+-----						
sigma_u		2.4062081				
sigma_e		.95849744				
rho		.86305286	(fraction of variance due to u_i)			
-----+-----						

```

. test ln_emp ln_cap

```

```

( 1) ln_emp = 0
( 2) ln_cap = 0

```

```

      F( 2, 1416) = 116.27
      Prob > F = 0.0000

```

```

.
. * -- G1. IV: the AI exposure of the firm's main occupation as an instrument f
> or ai_firm.

```

```
. * We run it to see what a doubtful instrument looks like. Only firm-years
> with ads have a
. * main occupation: a smaller, selected sample.
. keep if !missing(daioe_main)
(2,492 observations deleted)
```

```
. count
6,330
```

```
. * OLS on the same sample, for a like-with-like comparison
. regress ln_fin ai_firm ln_emp ln_cap i.year, vce(cluster firmID)
```

```
Linear regression          Number of obs   =      6,330
                          F(9, 1233)       =      86.70
                          Prob > F         =      0.0000
                          R-squared        =      0.3254
                          Root MSE      =      2.3822
```

(Std. err. adjusted for 1,234 clusters in firmID)

	ln_fin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ai_firm		.3084838	.1603996	1.92	0.055	-.0062027	.6231702
ln_emp		.3958199	.0743549	5.32	0.000	.2499439	.541696
ln_cap		.4939836	.0231397	21.35	0.000	.448586	.5393811
year							
2019		.0711369	.0572816	1.24	0.215	-.0412432	.1835171
2020		.1116417	.062065	1.80	0.072	-.0101231	.2334064
2021		.094028	.0641915	1.46	0.143	-.0319087	.2199647
2022		.1610611	.0656758	2.45	0.014	.0322124	.2899098
2023		.1252522	.0681068	1.84	0.066	-.0083659	.2588703
2024		.2369669	.0677591	3.50	0.000	.1040309	.3699028
_cons		-.2034058	.4252632	-0.48	0.633	-1.037725	.6309137

```
. * First stage first, then the reduced form, then 2SLS
. regress ai_firm daioe_main ln_emp ln_cap i.year, vce(cluster firmID)
```

```
Linear regression          Number of obs   =      6,330
                          F(9, 1233)       =      21.96
                          Prob > F         =      0.0000
                          R-squared        =      0.1025
                          Root MSE      =      .34508
```

(Std. err. adjusted for 1,234 clusters in firmID)

	ai_firm	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
daioe_main		.0017815	.0002332	7.64	0.000	.001324	.002239
ln_emp		.0958348	.0103761	9.24	0.000	.0754781	.1161915
ln_cap		-.0022793	.0030233	-0.75	0.451	-.0082106	.003652
year							
2019		.0409561	.0107984	3.79	0.000	.0197709	.0621413
2020		.0118455	.0114047	1.04	0.299	-.0105294	.0342203
2021		.0383601	.0111382	3.44	0.001	.0165082	.0602121
2022		.0689791	.0121833	5.66	0.000	.0450768	.0928814
2023		.0373974	.0116303	3.22	0.001	.0145801	.0602146
2024		.049132	.0120473	4.08	0.000	.0254966	.0727674
_cons		-.5431554	.0584712	-9.29	0.000	-.6578694	-.4284414

```
. test daioe_main
```

```
( 1) daioe_main = 0
```

```
F( 1, 1233) = 58.36
Prob > F = 0.0000
```

```
. regress ln_fin daioe_main ln_emp ln_cap i.year, vce(cluster firmID)
```

```
Linear regression               Number of obs   =      6,330
                                F(9, 1233)       =      88.69
                                Prob > F         =      0.0000
                                R-squared         =      0.3319
                                Root MSE      =      2.3707
```

(Std. err. adjusted for 1,234 clusters in firmID)

ln_fin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
daioe_main	.0065811	.0017189	3.83	0.000	.0032088	.0099534
ln_emp	.4298557	.0723408	5.94	0.000	.287931	.5717805
ln_cap	.4898459	.0230939	21.21	0.000	.4445382	.5351536
year						
2019	.0777036	.0568167	1.37	0.172	-.0337645	.1891717
2020	.1092615	.0620907	1.76	0.079	-.0125537	.2310767
2021	.0961084	.0644303	1.49	0.136	-.0302967	.2225136
2022	.1776087	.0646255	2.75	0.006	.0508205	.3043968
2023	.1324005	.0676746	1.96	0.051	-.0003697	.2651707
2024	.2483942	.0671704	3.70	0.000	.1166133	.3801752
_cons	-.6693302	.4226576	-1.58	0.114	-1.498538	.1598775

```
. * (with one instrument, the reduced form's test of daioe_main is the weak-
> IV-robust
. * test of "AI has no effect")
. ivregress 2sls ln_fin (ai_firm = daioe_main) ln_emp ln_cap i.year, vce(cluste
> r firmID)
```

```
Instrumental-variables 2SLS regression       Number of obs   =      6,330
                                                Wald chi2(9)     =     653.74
                                                Prob > chi2      =      0.0000
                                                R-squared        =      0.1565
                                                Root MSE        =      2.6617
```

(Std. err. adjusted for 1,234 clusters in firmID)

ln_fin	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ai_firm	3.694116	1.052626	3.51	0.000	1.631007	5.757225
ln_emp	.0758309	.1248472	0.61	0.544	-.1688651	.3205269
ln_cap	.4982659	.0252396	19.74	0.000	.4487972	.5477346
year						
2019	-.0735929	.081162	-0.91	0.365	-.2326674	.0854816
2020	.065503	.0754462	0.87	0.385	-.0823688	.2133747
2021	-.0455984	.0870968	-0.52	0.601	-.2163049	.1251081
2022	-.0772082	.1107734	-0.70	0.486	-.29432	.1399036
2023	-.0057497	.0919583	-0.06	0.950	-.1859846	.1744852
2024	.0668951	.0974683	0.69	0.493	-.1241393	.2579294
_cons	1.337149	.6442216	2.08	0.038	.0744977	2.5998

Endogenous: ai_firm

Exogenous: ln_emp ln_cap 2019.year 2020.year 2021.year 2022.year 2023.year
2024.year daioe_main

```
. display _b[ai_firm]
3.6941159
```

```
. estat firststage
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(1,1233)	Prob > F
----------	-------	-------------------	------------------	---------------------	----------

```

-----+-----
      ai_firm | 0.1025      0.1013      0.0393      58.3635      0.0000
-----+-----
(F statistic adjusted for 1234 clusters in firmID)

.
. * -- G2. Hausman said fixed effects, so a referee asks for IV with firm fixed
> effects too
. xtivreg ln_fin ln_emp ln_cap i.year (ai_firm = daioe_main), fe vce(cluster fi
> rmID) first

First-stage within regression

Fixed-effects (within) regression              Number of obs   =      6,330
Group variable: firmID                        Number of groups =      1,234

R-squared:                                    Obs per group:
  Within = 0.0161                               min =          1
  Between = 0.1141                             avg =          5.1
  Overall = 0.0646                             max =          7

                                          F(9, 1233)      =      7.89
corr(u_i, Xb) = 0.1798                      Prob > F         =      0.0000

```

(Std. err. adjusted for 1,234 clusters in firmID)

ai_firm	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ln_emp	.0335159	.0100558	3.33	0.001	.0137875	.0532442
ln_cap	-.0055891	.00702	-0.80	0.426	-.0193615	.0081833
year						
2019	.0461137	.010415	4.43	0.000	.0256805	.0665468
2020	.0247992	.0108103	2.29	0.022	.0035905	.0460079
2021	.0430295	.0107325	4.01	0.000	.0219735	.0640855
2022	.0780057	.0115046	6.78	0.000	.0554349	.1005765
2023	.0480458	.0111329	4.32	0.000	.0262042	.0698873
2024	.0583538	.0118329	4.93	0.000	.035139	.0815687
daioe_main	.0002453	.00028	0.88	0.381	-.0003041	.0007947
_cons	-.0492913	.0722965	-0.68	0.495	-.1911291	.0925464
sigma_u	.29357851					
sigma_e	.22505322					
rho	.62986004	(fraction of variance due to u_i)				

```

Fixed-effects (within) IV regression              Number of obs   =      6,330
Group variable: firmID                        Number of groups =      1,234

R-squared:                                    Obs per group:
  Within = .                                           min =          1
  Between = 0.1466                             avg =          5.1
  Overall = 0.1135                             max =          7

                                          Wald chi2(9)      =      252.52
corr(u_i, Xb) = 0.0395                      Prob > chi2       =      0.0000

```

(Std. err. adjusted for 1,234 clusters in firmID)

ln_fin	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ai_firm	-2.322376	6.661143	-0.35	0.727	-15.37798	10.73322
ln_emp	.4057601	.2262577	1.79	0.073	-.0376969	.8492171
ln_cap	.1987777	.0532737	3.73	0.000	.0943632	.3031921
year						
2019	.1845586	.311574	0.59	0.554	-.4261152	.7952324
2020	.1905578	.1740492	1.09	0.274	-.1505724	.5316881

```

2021 | .2572956 .2927705 0.88 0.379 -.316524 .8311151
2022 | .4393286 .5242166 0.84 0.402 -.5881171 1.466774
2023 | .4298801 .3272864 1.31 0.189 -.2115895 1.07135
2024 | .5252948 .3957725 1.33 0.184 -.2504052 1.300995
    |
_cons | 3.015447 .4293497 7.02 0.000 2.173937 3.856957
-----+-----
sigma_u | 2.5605537
sigma_e | 1.0874515
rho      | .84719565 (fraction of variance due to u_i)
-----+-----
Endogenous: ai_firm
Exogenous: ln_emp ln_cap 2019.year 2020.year 2021.year 2022.year 2023.year
           2024.year daioe_main

. display _b[ai_firm]
-2.3223761

. * STOP: relevance is tested; exclusion is an argument. Firms that hire for
> exposed
. * occupations may also finance themselves differently (leasing, debt), so
> exposure
. * can move financial expenses directly.
.
. log close
name: <unnamed>
log type: text
closed on: 16 Oct 2026, 17:06:02
-----+-----

```

6_report.log

```

name: <unnamed>
log type: text
opened on: 16 Oct 2026, 17:06:02

.
. * Writes one number as key,"value" to the numbers file
. capture program drop put

. program define put
1.     args key val
2.     file write nb ``key',"val'"" _n
3. end

. * A coefficient with stars as in the table from 4_regressions.do: *** 1%, **
> 5%, * 10%
. capture program drop star

. program define star, rclass
1.     args b p
2.     local s ""
3.     if `p' < 0.10 local s "*"
4.     if `p' < 0.05 local s "**"
5.     if `p' < 0.01 local s "***"
6.     return local txt = string(`b', "%9.3f") + "`s'"
7. end

.
. capture file close nb

. file open nb using "$root/output/tables/report_numbers.csv", write replace
(file
  ../project/output/tables/report_numbers.csv not found)

. file write nb "key,value" _n

.
. * -- R1. The sample, step by step (the same steps as 2_sample_vars.do)
. use "$root/data/processed/checked.dta", clear

```

(SYNTHETIC NA310G firm panel 2018-2024, dataset course – not real firms)

```
. count
  10,492

. local n0 = r(N)

. put n-row `': di %9.0fc `n0''

. egen byte t = tag(firmID)

. count if t
  1,534

. put firms-raw `': di %9.0fc r(N)'

. summarize year, meanonly

. local ymin = r(min)

. local ymax = r(max)

. put year-min `ymin'

. put year-max `ymax'

. drop if inlist(ser_laen, 8, 24)
(217 observations deleted)

. count
  10,275

. local n1 = r(N)

. put drop-counties `': di %9.0fc `n0' - `n1''

. put n-counties `': di %9.0fc `n1''

. gen finexp = -rr09_finkostn
(95 missing values generated)

. count if finexp == 0
  345

. put fin-zero `': di %9.0fc r(N)'

. count if finexp < 0
  0

. put fin-neg `': di %9.0fc r(N)'

. count if missing(finexp)
  95

. put fin-miss `': di %9.0fc r(N)'

. drop if missing(ln(finexp))
(440 observations deleted)

. count
  9,835

. local n2 = r(N)

. put drop-fin `': di %9.0fc `n1' - `n2''

. put n-fin `': di %9.0fc `n2''

. drop if missing(ln(bslov_antanst))
(268 observations deleted)

. count
  9,567
```

```

. local n3 = r(N)

. put drop-emp `: di %9.0fc `n2' - `n3''

. put n-emp `: di %9.0fc `n3''

. drop if missing(ln(br02a_byggmark + br03_maskiner))
(745 observations deleted)

. count
    8,822

. local n4 = r(N)

. put drop-cap `: di %9.0fc `n3' - `n4''

.
. file open tm using "$root/output/tables/table_sample.md", write replace
(file
  ../project/output/tables/table_sample.md not found)

. file write tm "| Step | Dropped | Firm-years left |" _n " |---|---:|---:" _n

. file write tm "| Course data, `ymin'-`ymax' | | `: di %9.0fc `n0'' |" _n

. file write tm "| Drop Kalmar and Västerbotten counties | `: di %9.0fc `n0'-`n
> 1'' | `: di %9.0fc `n1'' |" _n

. file write tm "| Drop: financial expenses zero, negative or missing | `: di %
> 9.0fc `n1'-`n2'' | `: di %9.0fc `n2'' |" _n

. file write tm "| Drop: no employees | `: di %9.0fc `n2'-`n3'' | `: di %9.0fc
> `n3'' |" _n

. file write tm "| Drop: no capital (buildings, land, machinery) | `: di %9.0fc
> `n3'-`n4'' | `: di %9.0fc `n4'' |" _n

. file close tm

.
. * -- R2. The estimation sample: size, AI shares, descriptive statistics, corr
> elations
. use "$root/data/processed/sample.dta", clear
(SYNTHETIC NA310G firm panel 2018-2024, dataset course - not real firms)

. xtset firmID year

Panel variable: firmID (unbalanced)
Time variable: year, 2018 to 2024, but with gaps
Delta: 1 unit

. count
    8,822

. put n-sample `: di %9.0fc r(N)'

. egen byte t = tag(firmID)

. count if t
    1,417

. put firms-sample `: di %9.0fc r(N)'

. summarize ai_firm, meanonly

. put ai-share `: di %4.1f 100*r(mean)'

. summarize ai_firm if year == 2018, meanonly

. put ai-share-first `: di %4.1f 100*r(mean)'

. summarize ai_firm if year == 2024, meanonly

```

```

. put ai-share-last `: di %4.1f 100*r(mean)'

. count if ai_firm == 0
7,827

. local z = r(N)

. count if ai_firm == 0 & has_ads == 0
2,492

. put zero-noads-pct `: di %4.0f 100*r(N)/`z''

. summarize finexp, detail

```

finexp				

Percentiles		Smallest		
1%	2	1		
5%	14	1		
10%	39	1	Obs	8,822
25%	232	1	Sum of wgt.	8,822
50%	1889.5		Mean	63605.92
		Largest	Std. dev.	521957.3
75%	12740	1.55e+07		
90%	65159	1.83e+07	Variance	2.72e+11
95%	188313	1.86e+07	Skewness	24.76503
99%	1112570	2.13e+07	Kurtosis	802.6558

```

. put fin-median `: di %9.0fc r(p50)'

. summarize ln_fin

```

Variable	Obs	Mean	Std. dev.	Min	Max
ln_fin	8,822	7.452756	2.874869	0	16.87222

```

. put sd-fin `: di %4.2f r(sd)'

.
. label variable ln_fin "Log financial expenses"

. label variable ln_emp "Log employees"

. label variable ln_cap "Log capital"

. label variable ai_firm "AI firm (0/1)"

. file open td using "$root/output/tables/table_desc.md", write replace
(file
.../project/output/tables/table_desc.md not found)

. file write td "| Variable | N | Mean | Median | SD | Min | Max |" _n " |---|---|
> -:|---:|---:|---:|---:|---:|" _n

. foreach v in ln_fin ai_firm ln_emp ln_cap {
2.   quietly summarize `v', detail
3.   file write td "| `: variable label `v'' | `: di %9.0fc r(N)' | `: di %
> 6.2f r(mean)' | `: di %6.2f r(p50)' | `: di %6.2f r(sd)' | `: di %6.2f r(min)
> ' | `: di %6.2f r(max)' |" _n
4. }

. file close td

. correlate ln_fin ai_firm ln_emp ln_cap
(obs=8,822)

```

	ln_fin	ai_firm	ln_emp	ln_cap
ln_fin	1.0000			
ai_firm	0.1047	1.0000		
ln_emp	0.3313	0.2408	1.0000	

```

ln_cap | 0.5485 0.0735 0.3657 1.0000

. matrix C = r(C)

. file open tc using "$root/output/tables/table_corr.md", write replace
(file
.../project/output/tables/table_corr.md not found)

. file write tc "| (1) | (2) | (3) | (4) |" _n " |---|:---|:---|:---|:---|
> " _n

. file write tc "| (1) `: variable label ln_fin' | 1 | | |" _n

. file write tc "| (2) `: variable label ai_firm' | `: di %4.2f C[2,1]' | 1 | |
> |" _n

. file write tc "| (3) `: variable label ln_emp' | `: di %4.2f C[3,1]' | `: di
> %4.2f C[3,2]' | 1 | |" _n

. file write tc "| (4) `: variable label ln_cap' | `: di %4.2f C[4,1]' | `: di
> %4.2f C[4,2]' | `: di %4.2f C[4,3]' | 1 |" _n

. file close tc

. put corr-fin-ai `: di %4.2f C[2,1]'

. put corr-fin-emp `: di %4.2f C[3,1]'

. put corr-fin-cap `: di %4.2f C[4,1]'

. put corr-ai-emp `: di %4.2f C[3,2]'

. put corr-emp-cap `: di %4.2f C[4,3]'

.
. * -- R3. The three models, clustered by firm (as in 4_regressions.do)
. regress ln_fin ai_firm ln_emp ln_cap i.year, vce(cluster firmID)

Linear regression                                Number of obs    =      8,822
                                                F(9, 1416)       =     108.13
                                                Prob > F         =      0.0000
                                                R-squared        =      0.3225
                                                Root MSE        =      2.3675

                                                (Std. err. adjusted for 1,417 clusters in firmID)

-----+-----
      ln_fin |               Robust
      Coefficient   std. err.      t    P>|t|     [95% conf. interval]
-----+-----
      ai_firm |      .3007456    .156736     1.92   0.055    - .0067141    .6082053
      ln_emp  |      .4210508    .0643936     6.54   0.000     .2947338    .5473679
      ln_cap  |      .4908312    .020946    23.43   0.000     .4497427    .5319197
      year   |
      2019    |      .0144356    .0402789     0.36   0.720    - .0645771    .0934483
      2020    |      .0814533    .0446086     1.83   0.068    - .0060527    .1689593
      2021    |      .0778108    .0483458     1.61   0.108    - .0170264    .1726479
      2022    |      .1698464    .0501316     3.39   0.001     .0715063    .2681865
      2023    |      .1937819    .0506637     3.82   0.000     .0943978    .2931659
      2024    |      .2298345    .051312     4.48   0.000     .1291788    .3304902
      _cons   |     - .3289511    .3694462    -0.89   0.373    -1.053672    .3957696
-----+-----

. put b-pooled `: di %5.3f _b[ai_firm]'

. put se-pooled `: di %5.3f _se[ai_firm]'

. put p-pooled `: di %5.3f 2*ttail(e(df_r), abs(_b[ai_firm]/_se[ai_firm]))'

. put pct-pooled `: di %4.1f 100*(exp(_b[ai_firm]) - 1)'

```

```

. put b-emp-pooled `: di %5.3f _b[ln_emp]'

. put b-cap-pooled `: di %5.3f _b[ln_cap]'

. put r2-pooled `: di %4.2f e(r2)'

. foreach v in ai_firm ln_emp ln_cap {
2.     star `=_b[\'v\']\' `=2*ttail(e(df_r), abs(_b[\'v\']/_se[\'v\']))'
3.     local P_\'v\' ""r(txt)""
4.     local Ps_\'v\' "(" + string(_se[\'v\'], "%9.3f") + ")"
5. }

. local P_N = e(N)

. local P_r2 = e(r2)

. regress ln_fin ai_firm ln_emp ln_cap i.year, vce(robust)

```

```

Linear regression                                Number of obs    =      8,822
                                                F(9, 8812)       =     506.10
                                                Prob > F         =     0.0000
                                                R-squared        =     0.3225
                                                Root MSE        =     2.3675

```

ln_fin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ai_firm	.3007456	.0845307	3.56	0.000	.1350456	.4664455
ln_emp	.4210508	.0292407	14.40	0.000	.3637322	.4783694
ln_cap	.4908312	.0089817	54.65	0.000	.4732249	.5084375
year						
2019	.0144356	.0935261	0.15	0.877	-.1688974	.1977685
2020	.0814533	.0938458	0.87	0.385	-.1025064	.2654129
2021	.0778108	.0940342	0.83	0.408	-.1065181	.2621397
2022	.1698464	.0944459	1.80	0.072	-.0152897	.3549824
2023	.1937819	.0932364	2.08	0.038	.0110168	.3765469
2024	.2298345	.0932973	2.46	0.014	.04695	.412719
_cons	-.3289511	.1771589	-1.86	0.063	-.6762238	.0183216

```

. put se-robust `: di %5.3f _se[ai_firm]'

. put p-robust `: di %5.3f 2*ttail(e(df_r), abs(_b[ai_firm]/_se[ai_firm]))'

. local pr = 2*ttail(e(df_r), abs(_b[ai_firm]/_se[ai_firm]))

. if `pr' < 0.001 put p-robust-txt "< 0.001"

. else put p-robust-txt `=' `: di %5.3f `pr''''

.
. xtreg ln_fin ai_firm ln_emp ln_cap i.year, fe vce(cluster firmID)

```

```

Fixed-effects (within) regression                Number of obs    =      8,822
Group variable: firmID                          Number of groups =      1,417

```

```

R-squared:                                     Obs per group:
  Within = 0.0657                               min =           1
  Between = 0.3434                               avg =          6.2
  Overall = 0.3047                               max =           7

```

```

corr(u_i, Xb) = 0.3535                        F(9, 1416)       =     46.19
                                                Prob > F         =     0.0000

```

(Std. err. adjusted for 1,417 clusters in firmID)

ln_fin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ai_firm	-.0517842	.0560016	-0.92	0.355	-.1616392	.0580708

ln_emp		.3462222	.032997	10.49	0.000	.2814941	.4109504
ln_cap		.1859054	.0263253	7.06	0.000	.1342646	.2375462
year							
2019		.0639915	.0361611	1.77	0.077	-.0069436	.1349267
2020		.1359775	.0385612	3.53	0.000	.0603344	.2116207
2021		.1628609	.0398493	4.09	0.000	.0846908	.2410309
2022		.2623581	.0403543	6.50	0.000	.1831975	.3415187
2023		.3237607	.0415182	7.80	0.000	.2423168	.4052046
2024		.3738758	.0420705	8.89	0.000	.2913487	.456403
_cons		3.232015	.2782346	11.62	0.000	2.686219	3.777811
sigma_u		2.4062081					
sigma_e		.95849744					
rho		.86305286	(fraction of variance due to u_i)				

```

. put b-fe `: di %5.3f _b[ai_firm]'

. put se-fe `: di %5.3f _se[ai_firm]'

. put p-fe `: di %5.3f 2*ttail(e(df_r), abs(_b[ai_firm]/_se[ai_firm]))'

. put pct-fe `: di %4.1f 100*(exp(_b[ai_firm]) - 1)'

. put lo-fe `: di %5.3f _b[ai_firm] - invttail(e(df_r), 0.025)*_se[ai_firm]'

. put hi-fe `: di %5.3f _b[ai_firm] + invttail(e(df_r), 0.025)*_se[ai_firm]'

. put b-emp-fe `: di %5.3f _b[ln_emp]'

. put p-emp-fe `: di %5.3f 2*ttail(e(df_r), abs(_b[ln_emp]/_se[ln_emp]))'

. put b-cap-fe `: di %5.3f _b[ln_cap]'

. put p-cap-fe `: di %5.3f 2*ttail(e(df_r), abs(_b[ln_cap]/_se[ln_cap]))'

. put r2w-fe `: di %4.2f e(r2_w)'

. put firms-fe `: di %9.0fc e(N_g)'

. foreach v in ai_firm ln_emp ln_cap {
2.     star `=_b[\'v\']' `=2*ttail(e(df_r), abs(_b[\'v\']/_se[\'v\'])))'
3.     local F_\'v' ``r(txt)''
4.     local Fs_\'v' = "(" + string(_se[\'v'], "%9.3f") + ")"
5. }

. local F_N = e(N)

. local F_r2 = e(r2_w)

. local F_g = e(N_g)

. test ln_emp ln_cap

( 1) ln_emp = 0
( 2) ln_cap = 0

      F( 2, 1416) = 116.27
      Prob > F = 0.0000

. put joint-F `: di %5.1f r(F)'

. put joint-p `: di %5.3f r(p)'

.

. bysort firmID: egen ever_ai = max(ai_firm)

. bysort firmID: egen always_ai = min(ai_firm)

. count if t & ever_ai == 1 & always_ai == 0
285

```

```

. put switchers `: di %9.0fc r(N)'

. count if t & always_ai == 1
51

. put always-ai `: di %9.0fc r(N)'

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

Random-effects GLS regression              Number of obs   =      8,822
Group variable: firmID                    Number of groups  =      1,417

R-squared:                                Obs per group:
    Within = 0.0617                        min =          1
    Between = 0.3543                       avg =         6.2
    Overall = 0.3208                       max =          7

corr(u_i, X) = 0 (assumed)                Wald chi2(9)      =     965.91
                                           Prob > chi2       =     0.0000

                                           (Std. err. adjusted for 1,417 clusters in firmID)
-----+-----
      ln_fin |      Coefficient      Robust      z      P>|z|      [95% conf. interval]
-----+-----
    ai_firm |    -.0162605    .0555056    -0.29    0.770    -.1250495    .0925285
    ln_emp  |     .3260244    .0295881    11.02    0.000     .2680328    .3840159
    ln_cap  |     .3626923    .0171697    21.12    0.000     .3290403    .3963442
    year    |
    2019    |     .0509667    .0361791     1.41    0.159    -.019943    .1218763
    2020    |     .1232491    .0387679     3.18    0.001     .0472655    .1992328
    2021    |     .1429643    .0400893     3.57    0.000     .0643907    .2215379
    2022    |     .2368771    .0404113     5.86    0.000     .1576724    .3160817
    2023    |     .2753829    .0411827     6.69    0.000     .1946662    .3560995
    2024    |     .3277374    .0421059     7.78    0.000     .2452114    .4102635
    _cons   |     1.532338    .221624     6.91    0.000     1.097963    1.966713
-----+-----
    sigma_u |  2.1731576
    sigma_e |  .95849744
    rho     |  .8371453   (fraction of variance due to u_i)
-----+-----

. put b-re `: di %5.3f _b[ai_firm]'

. put se-re `: di %5.3f _se[ai_firm]'

. put p-re `: di %5.3f 2*normal(-abs(_b[ai_firm]/_se[ai_firm]))'

. put r2o-re `: di %4.2f e(r2_o)'

. foreach v in ai_firm ln_emp ln_cap {
2.     star `=_b['v']' `=2*normal(-abs(_b['v']/_se['v']))'
3.     local R_`v' ``r(txt)''
4.     local Rs_`v' "(" + string(_se['v'], "%9.3f") + ")"
5. }

. local R_N = e(N)

. local R_r2 = e(r2_o)

.
. file open tr using "$root/output/tables/table_reg.md", write replace
(file
.../project/output/tables/table_reg.md not found)

. file write tr "| | (1) Pooled OLS | (2) Fixed effects | (3) Random effects |"
> _n "|---|:---:|:---:|:---:|" _n

. foreach v in ai_firm ln_emp ln_cap {

```

```

2.   file write tr "| `: variable label `v' | `P_`v' | `F_`v' | `R_`v'
> |" _n
3.   file write tr "| | `Ps_`v' | `Fs_`v' | `Rs_`v' |" _n
4. }

. file write tr "| Year effects | Yes | Yes | Yes |" _n " | Firm fixed effects |
> No | Yes | No |" _n

. file write tr "| Firm-years | `: di %9.0fc `P_N' | `: di %9.0fc `F_N' | `:
> di %9.0fc `R_N' |" _n

. file write tr "| Firms | `: di %9.0fc `F_g' | `: di %9.0fc `F_g' | `: di %9
> .0fc `F_g' |" _n

. file write tr "| R-squared | `: di %4.2f `P_r2' (overall) | `: di %4.2f `F_r
> 2' (within) | `: di %4.2f `R_r2' (overall) |" _n

. file close tr

.
. * -- R4. Which model: Hausman (default standard errors) and its robust versio
> n
. 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

```

	---- Coefficients ----		(b-B)	sqrt(diag(V_b-V_B))
	(b)	(B)	Difference	Std. err.
	fe_h	re_h		
ai_firm	-.0517842	-.0162605	-.0355237	.0117864
ln_emp	.3462222	.3260244	.0201979	.0147143
ln_cap	.1859054	.3626923	-.1767869	.0173291
year				
2019	.0639915	.0509667	.0130249	.0018474
2020	.1359775	.1232491	.0127284	.0020132
2021	.1628609	.1429643	.0198965	.002785
2022	.2623581	.2368771	.025481	.0033545
2023	.3237607	.2753829	.0483778	.0046852
2024	.3738758	.3277374	.0461384	.0046238

```

-----
b = Consistent under H0 and Ha; obtained from xtreg.
B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

      chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
              = 129.44
Prob > chi2 = 0.0000

. put hausman-chi2 `: di %5.1f r(chi2)'

. put hausman-df `r(df)'

. put hausman-p `: di %6.4f r(p)'

. bysort firmID: egen m_ai_firm = mean(ai_firm)

. bysort firmID: egen m_ln_emp = mean(ln_emp)

. bysort firmID: egen m_ln_cap = mean(ln_cap)

. forvalues y = 2019/2024 {
2.   bysort firmID: egen m_y`y' = mean(year == `y')
3. }

. xtreg ln_fin ai_firm ln_emp ln_cap m_ai_firm m_ln_emp m_ln_cap m_y* i.year, r

```

```
> e vce(cluster firmID)
```

```
Random-effects GLS regression      Number of obs   =      8,822
Group variable: firmID             Number of groups  =      1,417
```

```
R-squared:                          Obs per group:
    Within = 0.0657                      min =          1
    Between = 0.3583                     avg =          6.2
    Overall = 0.3268                     max =          7
```

```
corr(u_i, X) = 0 (assumed)           Wald chi2(18)    =    1300.68
                                      Prob > chi2       =      0.0000
```

(Std. err. adjusted for 1,417 clusters in firmID)

	ln_fin	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ai_firm		-.0517842	.0560302	-0.92	0.355	-.1616014	.058033
ln_emp		.3462222	.0330138	10.49	0.000	.2815163	.4109281
ln_cap		.1859054	.0263387	7.06	0.000	.1342824	.2375284
m_ai_firm		.5289002	.2457536	2.15	0.031	.047232	1.010568
m_ln_emp		.1048149	.0816343	1.28	0.199	-.0551853	.2648152
m_ln_cap		.3104624	.0336084	9.24	0.000	.2445912	.3763336
m_y2019		.1188686	1.376776	0.09	0.931	-2.579563	2.8173
m_y2020		-1.167846	1.164678	-1.00	0.316	-3.450573	1.114882
m_y2021		-.8367355	1.113546	-0.75	0.452	-3.019245	1.345774
m_y2022		.8953286	1.21383	0.74	0.461	-1.483735	3.274392
m_y2023		-.9786595	1.175671	-0.83	0.405	-3.282933	1.325614
m_y2024		-.5964793	.882118	-0.68	0.499	-2.325399	1.13244
year							
2019		.0639915	.0361796	1.77	0.077	-.0069192	.1349022
2020		.1359775	.0385809	3.52	0.000	.0603604	.2115946
2021		.1628609	.0398697	4.08	0.000	.0847177	.241004
2022		.2623581	.0403749	6.50	0.000	.1832247	.3414914
2023		.3237607	.0415395	7.79	0.000	.2423448	.4051765
2024		.3738758	.042092	8.88	0.000	.2913771	.4563746
_cons		-.321528	.7777685	-0.41	0.679	-1.845926	1.20287
sigma_u		2.1731576					
sigma_e		.95849744					
rho		.8371453	(fraction of variance due to u_i)				

```
. test m_ai_firm m_ln_emp m_ln_cap
```

```
( 1) m_ai_firm = 0
( 2) m_ln_emp = 0
( 3) m_ln_cap = 0
```

```
chi2( 3) = 120.46
Prob > chi2 = 0.0000
```

```
. put cre-chi2 `': di %5.1f r(chi2)'
```

```
. put cre-p `': di %6.4f r(p)'
```

```
. * -- R5. Tests for the preferred specification
. regress ln_fin ai_firm ln_emp ln_cap i.year
```

Source	SS	df	MS	Number of obs	=	8,822
Model	23513.6589	9	2612.62876	F(9, 8812)	=	466.13
Residual	49390.7663	8,812	5.60494397	Prob > F	=	0.0000
				R-squared	=	0.3225
				Adj R-squared	=	0.3218
Total	72904.4252	8,821	8.26487078	Root MSE	=	2.3675

ln_fin	Coefficient	Std. err.	t	P> t	[95% conf. interval]
--------	-------------	-----------	---	------	----------------------

ai_firm		.3007456	.0822276	3.66	0.000	.1395604	.4619308
ln_emp		.4210508	.0289166	14.56	0.000	.3643675	.4777341
ln_cap		.4908312	.0093699	52.38	0.000	.4724641	.5091983
year							
2019		.0144356	.0938755	0.15	0.878	-.1695824	.1984535
2020		.0814533	.0940161	0.87	0.386	-.1028402	.2657467
2021		.0778108	.0943769	0.82	0.410	-.10719	.2628115
2022		.1698464	.0944169	1.80	0.072	-.0152328	.3549256
2023		.1937819	.0940475	2.06	0.039	.0094267	.378137
2024		.2298345	.0937793	2.45	0.014	.0460053	.4136638
_cons		-.3289511	.1769863	-1.86	0.063	-.6758856	.0179834

```
. estat hettest, rhs iid-
```

BreuschPagan/-CookWeisberg test for heteroskedasticity

Assumption: i.i.d. error terms

Variables: All independent variables

H0: Constant variance

chi2(9) = 21.86

Prob > chi2 = 0.0093

```
. put het-chi2 `': di %5.1f r(chi2)'
```

```
. put het-p `': di %6.4f r(p)'
```

```
. estat vif
```

Variable		VIF	1/VIF
ai_firm		1.06	0.939042
ln_emp		1.22	0.818220
ln_cap		1.15	0.865844
year			
2019		1.70	0.587326
2020		1.70	0.589069
2021		1.69	0.590077
2022		1.70	0.588388
2023		1.70	0.587116
2024		1.71	0.583925
Mean VIF		1.52	

. * The VIFs estat vif prints above, from each regressor's auxiliary regression
> (year dummies included)

```
. quietly regress ai_firm ln_emp ln_cap i.year
```

```
. put vif-ai `': di %4.2f 1/(1 - e(r2))'
```

```
. quietly regress ln_emp ai_firm ln_cap i.year
```

```
. put vif-emp `': di %4.2f 1/(1 - e(r2))'
```

```
. quietly regress ln_cap ai_firm ln_emp i.year
```

```
. put vif-cap `': di %4.2f 1/(1 - e(r2))'
```

```
.
. * -- R6. The instrument considered in section 5 (as in 5_tests_iv.do)
. keep if !missing(daioe_main)
(2,492 observations deleted)
```

```
. count
6,330
```

```
. put n-iv `': di %9.0fc r(N)'
```

```
. regress ln_fin ai_firm ln_emp ln_cap i.year, vce(cluster firmID)
```

Linear regression

Number of obs	=	6,330
F(9, 1233)	=	86.70
Prob > F	=	0.0000
R-squared	=	0.3254
Root MSE	=	2.3822

(Std. err. adjusted for 1,234 clusters in firmID)

ln_fin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ai_firm	.3084838	.1603996	1.92	0.055	-.0062027	.6231702
ln_emp	.3958199	.0743549	5.32	0.000	.2499439	.541696
ln_cap	.4939836	.0231397	21.35	0.000	.448586	.5393811
year						
2019	.0711369	.0572816	1.24	0.215	-.0412432	.1835171
2020	.1116417	.062065	1.80	0.072	-.0101231	.2334064
2021	.094028	.0641915	1.46	0.143	-.0319087	.2199647
2022	.1610611	.0656758	2.45	0.014	.0322124	.2899098
2023	.1252522	.0681068	1.84	0.066	-.0083659	.2588703
2024	.2369669	.0677591	3.50	0.000	.1040309	.3699028
_cons	-.2034058	.4252632	-0.48	0.633	-1.037725	.6309137

```
. put b-ols-iv `: di %5.3f _b[ai_firm]'
```

```
. regress ai_firm daioe_main ln_emp ln_cap i.year, vce(cluster firmID)
```

Linear regression

Number of obs	=	6,330
F(9, 1233)	=	21.96
Prob > F	=	0.0000
R-squared	=	0.1025
Root MSE	=	.34508

(Std. err. adjusted for 1,234 clusters in firmID)

ai_firm	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
daioe_main	.0017815	.0002332	7.64	0.000	.001324	.002239
ln_emp	.0958348	.0103761	9.24	0.000	.0754781	.1161915
ln_cap	-.0022793	.0030233	-0.75	0.451	-.0082106	.003652
year						
2019	.0409561	.0107984	3.79	0.000	.0197709	.0621413
2020	.0118455	.0114047	1.04	0.299	-.0105294	.0342203
2021	.0383601	.0111382	3.44	0.001	.0165082	.0602121
2022	.0689791	.0121833	5.66	0.000	.0450768	.0928814
2023	.0373974	.0116303	3.22	0.001	.0145801	.0602146
2024	.049132	.0120473	4.08	0.000	.0254966	.0727674
_cons	-.5431554	.0584712	-9.29	0.000	-.6578694	-.4284414

```
. put fs-b `: di %6.4f _b[daioe_main]'
```

```
. test daioe_main
```

```
( 1) daioe_main = 0
```

```
F( 1, 1233) = 58.36
Prob > F = 0.0000
```

```
. put fs-F `: di %5.1f r(F)'
```

```
. regress ln_fin daioe_main ln_emp ln_cap i.year, vce(cluster firmID)
```

Linear regression

Number of obs	=	6,330
F(9, 1233)	=	88.69
Prob > F	=	0.0000
R-squared	=	0.3319

Root MSE = 2.3707

(Std. err. adjusted for 1,234 clusters in firmID)

ln_fin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
daioe_main	.0065811	.0017189	3.83	0.000	.0032088	.0099534
ln_emp	.4298557	.0723408	5.94	0.000	.287931	.5717805
ln_cap	.4898459	.0230939	21.21	0.000	.4445382	.5351536
year						
2019	.0777036	.0568167	1.37	0.172	-.0337645	.1891717
2020	.1092615	.0620907	1.76	0.079	-.0125537	.2310767
2021	.0961084	.0644303	1.49	0.136	-.0302967	.2225136
2022	.1776087	.0646255	2.75	0.006	.0508205	.3043968
2023	.1324005	.0676746	1.96	0.051	-.0003697	.2651707
2024	.2483942	.0671704	3.70	0.000	.1166133	.3801752
_cons	-.6693302	.4226576	-1.58	0.114	-1.498538	.1598775

```
. put rf-b `: di %6.4f _b[daioe_main]'
```

```
. put rf-p `: di %6.4f 2*ttail(e(df_r), abs(_b[daioe_main]/_se[daioe_main]))'
```

```
. ivregress 2sls ln_fin (ai_firm = daioe_main) ln_emp ln_cap i.year, vce(cluster
```

```
> r firmID)
```

Instrumental-variables 2SLS regression

Number of obs = 6,330

Wald chi2(9) = 653.74

Prob > chi2 = 0.0000

R-squared = 0.1565

Root MSE = 2.6617

(Std. err. adjusted for 1,234 clusters in firmID)

ln_fin	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ai_firm	3.694116	1.052626	3.51	0.000	1.631007	5.757225
ln_emp	.0758309	.1248472	0.61	0.544	-.1688651	.3205269
ln_cap	.4982659	.0252396	19.74	0.000	.4487972	.5477346
year						
2019	-.0735929	.081162	-0.91	0.365	-.2326674	.0854816
2020	.065503	.0754462	0.87	0.385	-.0823688	.2133747
2021	-.0455984	.0870968	-0.52	0.601	-.2163049	.1251081
2022	-.0772082	.1107734	-0.70	0.486	-.29432	.1399036
2023	-.0057497	.0919583	-0.06	0.950	-.1859846	.1744852
2024	.0668951	.0974683	0.69	0.493	-.1241393	.2579294
_cons	1.337149	.6442216	2.08	0.038	.0744977	2.5998

Endogenous: ai_firm

Exogenous: ln_emp ln_cap 2019.year 2020.year 2021.year 2022.year 2023.year

2024.year daioe_main

```
. put b-iv `: di %5.2f _b[ai_firm]'
```

```
. put se-iv `: di %5.2f _se[ai_firm]'
```

```
. put lo-iv `: di %5.2f _b[ai_firm] - invnormal(0.975)*_se[ai_firm]'
```

```
. put hi-iv `: di %5.2f _b[ai_firm] + invnormal(0.975)*_se[ai_firm]'
```

```
. xtreg ai_firm daioe_main ln_emp ln_cap i.year, fe vce(cluster firmID)
```

Fixed-effects (within) regression

Group variable: firmID

Number of obs = 6,330

Number of groups = 1,234

R-squared:

Within = 0.0161

Obs per group:

min = 1

```

Between = 0.1141          avg =      5.1
Overall = 0.0646          max =      7

corr(u_i, Xb) = 0.1798    F(9, 1233) =      7.89
                          Prob > F    =      0.0000

```

(Std. err. adjusted for 1,234 clusters in firmID)

ai_firm	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
daioe_main	.0002453	.00028	0.88	0.381	-.0003041	.0007947
ln_emp	.0335159	.0100558	3.33	0.001	.0137875	.0532442
ln_cap	-.0055891	.00702	-0.80	0.426	-.0193615	.0081833
year						
2019	.0461137	.010415	4.43	0.000	.0256805	.0665468
2020	.0247992	.0108103	2.29	0.022	.0035905	.0460079
2021	.0430295	.0107325	4.01	0.000	.0219735	.0640855
2022	.0780057	.0115046	6.78	0.000	.0554349	.1005765
2023	.0480458	.0111329	4.32	0.000	.0262042	.0698873
2024	.0583538	.0118329	4.93	0.000	.035139	.0815687
_cons	-.0492913	.0722965	-0.68	0.495	-.1911291	.0925464
sigma_u	.29357851					
sigma_e	.22505322					
rho	.62986004	(fraction of variance due to u_i)				

```

. put fefs-b `: di %6.4f _b[daioe_main]'

. put fefs-p `: di %5.3f 2*ttail(e(df_r), abs(_b[daioe_main]/_se[daioe_main]))'

.
. file close nb

. log close
  name: <unnamed>
  log type: text
  closed on: 16 Oct 2026, 17:06:05

```