Applied Econometrics

Testing quantitative empirical models

Frank Moers

Maastricht University

Intro

- The key to causal inference is control for observed confounding factors
  - What if there are unobserved confounding factors (unobserved heterogeneity?

Use panel data structure
(this session)

Instrumental Variables
(next session)

Partial identification
(last session)

Panel data

- Data with both a cross-sectional component (*i*) and a time component (*t*)
- Let's assume that

$$\{y_{0it}, y_{1it}\} \quad E[y_{0it}|A_i, X_{it}, t, D_{it}] = E[y_{0it}|A_i, X_{it}, t\ ]$$

- Also assume that A_i is not observed, which creates a problem in identifying the causal effect of D_{it}
  - (bc then the potential outcome is not independent of treatment and CIA is violated)

- Because A_i does not have a time subscript, it implies that it is a time-invariant unobserved confounding factor
  - If so, fixed effects has the potential to address the issue of identification

Fixed effects and its assumptions

- Assumption 1: linear model, i.e.,

$$E[y_{0it}|A_i, X_{it}, t] = \alpha + \lambda_t + \gamma A_i + \delta X_{it}$$

- Assumption 2: the causal effect is additive, i.e.,

$$E[y_{1it}|A_i, X_{it}, t] = E[y_{0it}|A_i, X_{it}, t] + \rho$$

$$\Downarrow$$

$$y_{it} = \alpha_i + \lambda_t + \rho D_{it} + \delta X_{it} + \varepsilon_{it}$$

where $\alpha_i = \alpha + \gamma A_i$

α_i are the "individual" fixed effects

λ_t are the "year" fixed effects

Estimation of fixed effects

- Option 1: Calculate all (!) variables as deviations from *individual-specific* means and use these variables in a regular OLS regression, i.e.,

$$(y_{it} - \bar{y}_i) = \left(\lambda_t - \bar{\lambda}\right) + \rho(D_{it} - \bar{D}_i) + \delta(X_{it} - \bar{X}_i) + \nu_{it}$$

  - The unobserved individual effects are "killed"

- Option 2: Calculate first-differences for all (!) variables and use these variables in a regular OLS regression, i.e.,

$$\Delta y_{it} = \Delta\lambda_t + \rho\Delta D_{it} + \delta\Delta X_{it} + \mu_{it}$$

  - The unobserved individual effects are again "killed"

- Option 1 = Option 2 when there are only two time periods

Problem with fixed effects

- Persistence in the treatment variable
  - $Var(D_{it} - \bar{D}_i)$ might be too low and the fixed effects have killed both bad and good variation
    - Let's assume the extreme situation where the treatment is constant *within* individuals. Then $D_{it} = \bar{D}_i \;\; \forall\, i$ and thus $Var(D_{it} - \bar{D}_i) = Var(\Delta D_{it}) = 0$. There is nothing left to estimate

- Measurement error in treatment variable might not be constant across time → $D_{it} - \bar{D}_i$ captures (also) noise

- Both problems create an attenuation bias (bias towards zero)

Associations and causality

- How can you deal with selection bias?
  - Panel-data techniques
    - Fixed effects (firm, individual, etc.)
    - Assumptions:
      - Unobserved confounding factors are time-invariant, i.e., constant over time
      - Model is a linear, additive model

    - Good for (credible) associations, but typically insufficient for causal statements

  - If the assumption of time-invariant unobserved confounding factors is not plausible, you are better of using the lagged dependent variable as control instead of using fixed effects (more later)

More "fixed effects"

- What if treatment changes for some individuals at a specific point in time and is constant before and after this change for these individuals?
  - That is, $D_{it} = 0 \;\; \exists\, i \;\; \forall\, t$ (for some *i* for all t) and $\{D_{it} = 0 | t < \hat{t}; D_{it} = 1 | t \geq \hat{t}\} \;\; \exists\, i$
- We can now try to use this "shock" to the treatment to identify the causal effect of D_{it}

$$y_{it} = \alpha_i + \lambda_t + \rho D_{it} + \varepsilon_{it}$$

  - Same as before but now ρ is the difference-in-differences estimator

$$\left.\begin{aligned} \Delta_0 &= E[y_{0it}|\alpha_i, t \geq \hat{t}] - E[y_{0it}|\alpha_i, t < \hat{t}] = \lambda_{t \geq \hat{t}} - \lambda_{t < \hat{t}} \\ \Delta_1 &= E[y_{1it}|\alpha_i, t \geq \hat{t}] - E[y_{1it}|\alpha_i, t < \hat{t}] = \lambda_{t \geq \hat{t}} - \lambda_{t < \hat{t}} + \rho \end{aligned}\right\} \Delta_1 - \Delta_0 = \rho$$

Difference-in-Differences (DiD)

- The typical DiD analysis, however, involves more “aggregate” data, i.e., pre-specified “groups” (do not) get exposed to the treatment → group *s*

$$y_{ist} = \gamma_s + \lambda_t + \rho D_{st} + \varepsilon_{ist}$$

  - Note that “*i*-fixed effects” are not necessary here for a consistent estimate of ρ

- A regression version of the above equation allows you to, for example,
  - Add time-variant covariates at the *s*-level (adding them at the *i*-level only affects precision)
  - Test for Granger causality

DiD: an example

- Does patenting decrease as insider trading opportunities decrease?
  - Insider trading opportunities allow executives to extract rents from innovations by trading before the patent grant is publicly known → incentives to patent
  - The American Inventors Protection Act (AIPA) of 1999 limited insider trading opportunities by introducing pre-grant disclosure of patent applications
  - Only executives of listed firms (that patent) are exposed to the insider trading effects of the AIPA
    - Executives of private firms (that patent) are by construction unaffected (in terms of insider trading)

- Thus
  - *i*: individual patenting firms, both public and private
  - *s*: public firm versus private firm
  - *t*: year, with $\hat{t} = 1999$
  - $[D_{st} | s = private] = 0$
  - $[D_{st} | s = public, t < \hat{t}] = 0$
  - $[D_{st} | s = public, t > \hat{t}] = 1$

$$y_{ist} = \gamma_s + \lambda_t + \rho D_{st} + \varepsilon_{ist}$$

DiD: an example

$$y_{ist} = \gamma_s + \lambda_t + \rho D_{st} + \varepsilon_{ist}$$

- *Patenting*: (citation-weighted) granted patent applications
- *Listed*: 1 for public firms, 0 for private firms
- *PostAIPA*: 1 for years after 1999, 0 for years before 1999 (1999 is dropped)

- Basic regression specification:

$$Patenting_{ist} = \beta_0 + \beta_1 Listed_s + \beta_2 Listed_s \cdot PostAIPA_t + \lambda_t + \varepsilon_{ist}$$

- $\beta_0 + \beta_1 Listed_s \equiv \gamma_s$
- $Listed_s \cdot PostAIPA_t \equiv D_{st} \rightarrow \beta_2$ is the estimate of ρ
- Note that the main effect of $PostAIPA_t$ is kicked out
    - It is a perfect linear combination of the year fixed effects $\lambda_t \rightarrow PostAIPA_t = (I|t > \hat{t})\lambda_t$

DiD: an example

- Actual regression specifications used

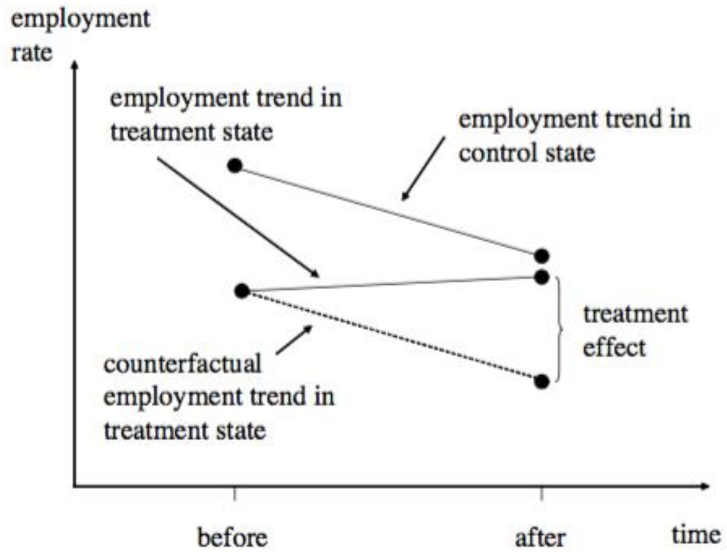
$$
\begin{aligned}
Patenting_{ist} &= \beta_0 + \beta_1 Listed_s + \beta_2 Listed_s \cdot PostAIPA_t + \lambda PostAIPA_t + \delta X_{ist} + \varepsilon_{ist} \\
Patenting_{ist} &= \beta_0 + \beta_1 Listed_s + \beta_2 Listed_s \cdot PostAIPA_t + \lambda_t + \delta X_{ist} + \varepsilon_{ist} \\
Patenting_{ist} &= \beta_0 + \alpha_i + \beta_2 Listed_s \cdot PostAIPA_t + \lambda_t + \delta X_{ist} + \varepsilon_{ist}
\end{aligned}
$$

  - added time-variant covariates at the firm level (X_{ist}) as well as firm fixed effects (α_i) are only added for efficiency reasons

	(1)	(2)	(3)	(4)	(5)	(6)
		# Patents			Class & Year-Adj. # Citations	
Post AIPA	0.03***			0.01		
	(4.75)			(0.73)		
Listed	0.40***	0.40***		0.46***	0.46***	
	(18.29)	(18.13)		(15.42)	(15.36)	
Post AIPA · Listed	-0.21***	-0.19***	-0.12***	-0.20***	-0.20***	-0.09**
	(-8.05)	(-7.37)	(-3.33)	(-5.08)	(-5.04)	(-1.98)
Ln(# Patents last 5 years)	0.70***	0.70***	-0.09***	0.72***	0.72***	-0.09***
	(135.34)	(135.17)	(-6.61)	(94.44)	(94.52)	(-5.93)
Ln(# Years since first patent)	-0.38***	-0.38***	0.27***	-0.46***	-0.46***	0.13***
	(-80.01)	(-80.25)	(11.65)	(-57.30)	(-57.27)	(4.30)
Year Fixed Effects	No	Yes	Yes	No	Yes	Yes
Firm Fixed Effects	No	No	Yes	No	No	Yes
Observations	151,220	151,220	120,536	151,220	151,220	107,698

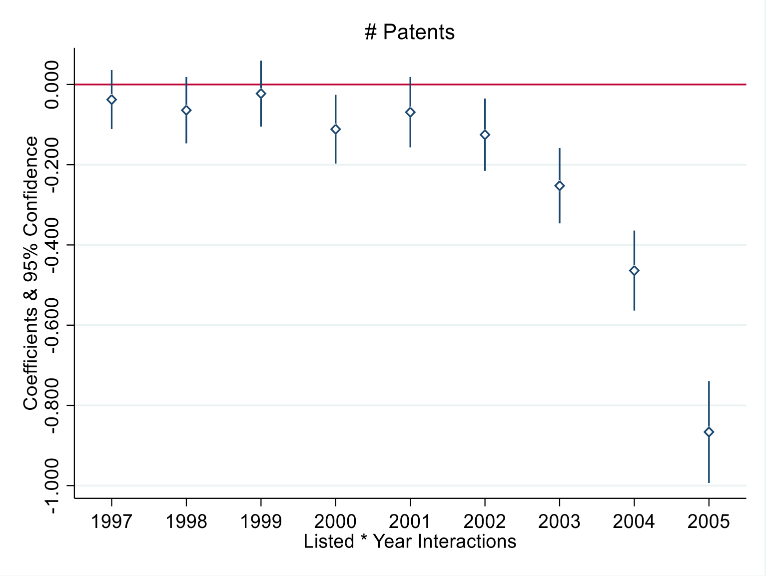
Associations and causality

- How can you deal with selection bias?
  - Panel-data techniques
    - Difference-in-differences
      - Pre and post; treatment and control



DiD: an example

- The parallel trend assumption is, by construction, an assumption
  - But you can test to what extent this assumption is plausible by
    - Showing the trends of both groups in the pre-treatment period, and/or
    - Show that the parallel trend "breaks" upon the treatment kicking in
      - Run a regression with year fixed effects (λ_t) interacted with the group variable (*Listed*) and "show" interactions



Associations and causality

- How can you deal with selection bias?
  - Panel-data techniques
    - Difference-in-differences
      - Pre and post; treatment and control
    - Key assumption: parallel trends
      - Treatment would have the same trend as the control if it would not have been treated
    - Has the potential for causal statements!

One final note

- If there are both time-invariant and time-variant unobserved confounding factors, why not use both fixed effects and the lagged dependent variable?
  - Without any adjustments, such a setup will have, by design, a correlation between the lagged DV and the residual → OLS estimates are not consistent and can thus not be used

- Dynamic panel data model
  - Using both the lagged DV and fixed effects requires dynamic panel data modeling
  - If interested, the following paper is a good read (also if you do not use Stata):
    - Roodman, D. 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal* 9: 86-136.