Tutorial, Part 15

Panel Data (static)

Example 1

Panel data : cross-section (i) & time-series (t)

POOLED (stata: xtset i-var t-var)

Balanced panel: all units (i) have the same number of time observations (t)

X_{it} not independent

With e.g. OLS estimation we neglect the panel structure of the data → bias, inconsistency

↳ Fixed-effects estimator, Random-effects estimator

a) 10 firms, 15 years

Pattern of an unbalanced panel: the units usually drop out. Bottom row: x - if at least one firm observed in that year

Source of variation: - overall : $X_{it} - \bar{\bar{X}}$ ← overall average

(i - one unit, „group") - between : $\bar{X}_i - \bar{\bar{X}}$

- within : $X_{it} - \bar{X}_i$ ← unit (group) i average

b) LSDV (Least Squares Dummy Variable) Estimator

$$Q_{it} = \beta_0 F_{it}^{\beta_1} M_{it}^{\beta_2} E_{it}^{\beta_3} L_{it}^{\beta_4} P_{it}^{\beta_5} \cdot u_{it} \qquad / \log$$

$$\log Q_{it} = \log\beta_0 + \beta_1 \log F_{it} + \beta_2 \log M_{it} + \beta_3 \log E_{it} + \beta_4 \log L_{it} + \beta_5 \log P_{it} + u_{it} + c_i$$

FEE: $c_i \rightarrow \alpha_i$ (fixed) $\Rightarrow E(\alpha_i | x_i) = g(x_i) \neq 0$ (no general intercept)

REE: $c_i \rightarrow v_i$ (variable, random) $\Rightarrow E(v_i | x_i) = 0$

↳ $u_{it} + v_i = \omega_{it}$ (compound disturbance)

Dummy coefficients: differential intercept coefficients
$\hat{\alpha}_i = \hat{\alpha}_1 + \text{coeff}_i \Rightarrow$ fixed = constant over time

LSDVE → add dummies → OK for small n

Example: partial elasticities (interpretation!)

FEE (Fixed Effects Estimator) (the „within" regression)

$n = 125$
df total $= n - 1 = 124$
m_1 = # „true" explanatory vars = 5
$m_2 = (n-1) - m_1 -$ # dummies $= 110$
! corr ($\overset{\alpha_i}{u_i}$, xb) $\neq 0$
⇒ same coeffs for main expl. vars as with LSDVE
⇒ _cons $= \overline{\alpha_i}$ (mean FE)

F-test: $H_0: \alpha_i = 0 \ \forall i$ df: # dummies
$H_1: \alpha_i \neq 0 \ \exists i$

⇒ „manual" calculation of FE ($\hat{\alpha}_i$)
$\hat{\alpha}_i = \overline{Y}_{i.} - X_{i.}^T \hat{\beta}_{FE}$; $i = 1, 2, \ldots, n$ (10)
vectors of means

But: FEE
- neglects between variation
- loss of df due to too many dummies
- elimination of time-invariant expl. vars.
(due to the „within" transformation $y_{it} - \bar{y}_i = \tilde{y}_{it}$)

⇒ REE
(preferable, if it passes the Hausman test)

c) REE (Random Effects Estimator)

(weighted average of within and between estimator)

→ individual specific constant terms considered as randomly distributed ($\alpha_i \to \nu_i$)

- the disturbance term thus not estimable
- the disturbance term uncorrelated with regressors ($E(\nu_i | x_i) = 0$)
- the model includes intercept

REE :
- more efficient
- better prediction
- less parameters
- time-invariant expl. vars not neglected

d) But REE can be used only if it „passes" the Hausman test:

H_0: FEE & REE consistent, FEE inefficient → use REE

H_1: REE inconsistent → use FEE

Estimator (model) \ True process	RE (H_0 not rejected)	FE (H_0 rejected)
REE	· consistent · efficient	INCONSISTENT! (efficiency doesn't matter)
FEE	· consistent · inefficient	· consistent · possibly efficient

if not sure: use FEE

e) reject H_0 → use FEE

Constant returns to scale (Σ elasticities $= 1$)

$H_0: \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 = 1$

$H_1: \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 \neq 1$ ⟹ cannot reject H_0

Example 2

a) elasticities

$H_0: \alpha_i = 0 \quad \forall i \quad \Rightarrow$ reject H_0

b) year-dummies not sig. anymore as wage vars. explain changes through time

H_0: all these coeffs. $= 0$

H_1: at least one coeff. $\neq 0$ $\quad \Rightarrow$ reject H_0

c) First-difference OLS (3rd option for FE)

$$\Delta y_{it} = \Delta x_{it}^T \beta + \Delta u_{it}$$

$\rightarrow$ usually less efficient (higher standard errors)

d) Hausman test

H_0: REE

H_1: FEE $\quad \Rightarrow$ reject H_0 $\rightarrow$ use FEE

e) H_0: homoscedasticity (stata: xttest3)

H_1: heteroscedasticity $\quad \Rightarrow$ reject H_0

H_0: no first-order autocorrelation (stata: xtserial)

H_1: first-order autocorrelation present $\quad \Rightarrow$ reject H_0

Solution: xtregar ..., fe (in case of autocorrelation)

xtreg ..., fe vce(robust) (in case of heteroscedasticity)

Tutorial, Part 16

Panel data (dynamic)

Example 1

a) • Static panel data model : long-run effects
• Dynamic panel data model : short-run effects
↳ past values of the expl. vars affect the current value of the dep. var (through y_{t-1})

! Dynamic panel data estimators are highly sensitive to particular specification of the model and the instruments.
→ robustness checks → experimentation

b) Arellano-Bond estimator : (GMM* estimator)
„internal" and „external" instruments →
more efficient (compared to Anderson-Hsiao ← only „internal" i.v.)
↳ substantial loss of df

---

* GMM - another approach to estimation:
finite-dimensional parameters but unknown shape of distribution → MLE not possible
Generalized Method of Moments (1st Moment: mean, 2nd: variance, 3rd: skewness..)
↳ „powers" of deviations

---

stata : xtabond 2
(gmmstyle → lags of dependent variable
ivstyle → exogenous variables)

-1* → inclusion of time dummies to prevent contemporaneous („cross-individual") correlation
twostep → VCE matrix estimated by levels and differences

$\hat{\rho}$ - the coeff. with lagged dep. var. ($y_{i,t-1}$)
$\hat{\rho} < 1$ $\longrightarrow$ dynamic stability

Diagnostics:

- autocorrelation: • of order 1 → present by construction
  • of higher orders → shouldn't be present
- joint validity of moment conditions:
  (= identifying restrictions = joint validity of instruments)
  • Sargan test: not robust, not weakened by many instruments
  • Hansen J-test: robust, weakened by many instruments → nevertheless superior to Sargan
    → H_0: instruments are valid

c) improvement: instruments are lagged levels & lagged differences (lagged levels poor instruments for Δ)
⇒ system of two equations:
- first differences equation
- levels equation

} system GMM

(stata: remove noleveleq option)

(Z matrix)

d) Collapse the columns of all available lags as instruments into a single column (1 instrument for var & lag, not 1 for var & lag & time observation)
→ recommended for small samples (avoids bias due to too many instruments)

e) 3 ways of dealing with endogeneity in the model:
- „within" transformation: subtracts the average of all values from the current observation
- first-difference (FD, Δ) transformation (GMM default): subtracts the previous value from the current observation (→ drops 1st observation, magnifies gaps in an unbalanced p.)
- forward orthogonal deviations (FOD) transformation: subtracts the average of all available future observations from the current value (→ drops last observation)

f) Coeffs from dynamic panel data models are estimates of short-run effects.
Long-run effect: $b_{j,long} = b_{j,short} \cdot \frac{1}{1-\hat{\rho}}$
(Stata: nlcom)
→ Interpretation: as in an ordinary regression function