Panel Data Analysis

Statistics and Econometrics

Jiahua Wu

382 Business School
j.wu@imperial.ac.uk

Roadmap

- Regression analysis with cross-sectional data
  - Basics: estimation, inference, analysis with dummy variables
  - More involved: model specification and data issues
- Advanced topics
  - Binary dependent variable models
  - Panel data analysis
  - Time series analysis

Outline (Wooldridge, Chap. 13.3, 13.5, 14.1)

- Fixed effects panel model
- Estimation of fixed effects model
    - First differencing
    - Fixed effects

Outline

- Fixed effects panel model
- Estimation of fixed effects model
  - First differencing
  - Fixed effects

What is Panel Data?

- A set of panel data
  - has both a cross-sectional and a time series dimension
  - is collected by following the same individuals over a number of time periods
- Panel data allows us to address issues related to unobserved factors, which are difficult to handle with cross sectional data

Two-Period Panel Data

- Example 9.4. City Crime Rates
  - Data: crime rates (*crmrte*) and unemployment rates (*unem*) from a sample of 46 cities in 1982 ($t = 1$) and 1987 ($t = 2$).
  - Question: Did *unem* influence *crmrte*?
  - Regressing *crmrte* on *unem* using the sample from 1987, we have

$$\widehat{crmrte}_{87} = \underset{(20.76)}{128.38} - \underset{(3.42)}{4.16}\, unem_{87},$$

    $n = 46, R^2 = .033$
  - The result is likely biased because many relevant factors (e.g., city, police, ...) are not controlled for

Two-Period Panel Data

- An alternative way to look at the data
  - If the omitted variables are fixed over time, then we can decompose the error into two parts: factors that vary over time and those do not
- Consider the previous example in the panel setting

$$crmrte_{it} = \beta_0 + \delta_0 d2_t + \beta_1 unem_{it} + a_i + u_{it}, \quad t = 1, 2$$

where

  - i is the city
  - t is the time period
  - $d2_t$ is the dummy variable indicating the second time period
  - A time-constant component is added to the error $v_{it} = a_i + u_{it}$

Fixed Effects Model

- In general, the fixed-effects model can be written as

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it}, \quad t = 1, 2$$

  where

  - a_i is the fixed effect (invariant to t) that represents factors specific to individual i (allowed to be correlated with $\mathbf{x}_{it}$)
  - u_{it} is called the idiosyncratic error that represents unobserved factors varying both overtime and across sections (typically assumed to be uncorrelated with $\mathbf{x}_{it}$)

Outline

- Fixed effects panel model
- Estimation of fixed effects model
    - First differencing
    - Fixed effects

First Differencing

- Write the model separately

$$y_{i1} = \beta_0 + \delta_0 \cdot 0 + \beta_1 x_{i11} + \cdots + \beta_k x_{i1k} + a_i + u_{i1}, \quad (t = 1)$$
$$y_{i2} = \beta_0 + \delta_0 \cdot 1 + \beta_1 x_{i21} + \cdots + \beta_k x_{i2k} + a_i + u_{i2}, \quad (t = 2)$$

- Subtracting the first equation from the second one gives

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_{i1} + \cdots + \beta_k \Delta x_{ik} + \Delta u_i,$$

(first-differenced equation) which is a cross-section model and is free of a_i

Panel Data Estimation in `R`

- The command to perform panel data estimation in `R` is `plm(formula, data, effect, model, index, ...)`
  - `effect`
    - fixed effects for cross-sectional units ("`individual`")
  - `model`
    - first-differences ("`fd`")
    - fixed effects ("`within`")

First Differencing: An Example

- Example 9.7. City Crime Rates.
  - First differencing

$$\Delta\widehat{crmrte} = \underset{(4.70)}{15.40} + \underset{(.88)}{2.22}\Delta unem,$$

  $n = 46, R^2 = .127$

    - There is a positive and significant relationship between $unem_{it}$ and $crmrte_{it}$
    - One percentage point rise in unemployment rate increases 2.22 crimes per 1,000 people
    - The crimes per 1,000 people increased by 15.4 in 1987, in comparison to 1982

Panel Data with More than Two Periods

- For the panel data with T periods
  1. Subtract period 1 from period 2

     $\vdots$

  2. Subtract period $(T - 1)$ from period T
  3. We have $(T - 1)$ observations per individual
  4. Estimate by OLS

Outline

- Fixed effects panel model
- Estimation of fixed effects model
  - First differencing
  - Fixed effects

Fixed Effects Estimation

- When there is an unobserved fixed effect, an alternative to first differencing is fixed effects estimation
- Consider a model with a single independent variable

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

  - The average over time for individual i is

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_i$$

  - The average of a_i will be a_i. So if we subtract the average from y_{it}, we have

$$y_{it} - \bar{y}_i = \beta_1 (x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i)$$

  - Each individual has been "de-meaned" for all variables, which eliminates the fixed effects.

Example 9.7. City Crime Rates

```
> crime.fd <- plm(crmrte ~ d87 + unem, data, index = c("city", "year"),
                  effect = "individual", model = "fd")
> crime.fe <- plm(crmrte ~ d87 + unem, data, index = c("city", "year"),
                  effect = "individual", model = "within")
> stargazer(crime.fd, crime.fe, header = FALSE, type = 'latex',
          column.labels = c("fd","fe"))
```

	Dependent variable:	
	crmrte	
	fd	fe
	(1)	(2)
d87		15.402***
		(4.702)
unem	2.218**	2.218**
	(0.878)	(0.878)
Constant	15.402***	
	(4.702)	
Observations	46	92
R^2	0.127	0.196
Adjusted R^2	0.107	−0.663
F Statistic	6.384** (df = 1; 44)	5.365*** (df = 2; 44)
Note:		*p<0.1; **p<0.05; ***p<0.01