Panel Models: Theoretical Insights

David Bell
University of Stirling

Lecture Structure

- Rationale for Panel Models
- Construction of one-way and two-way error components models
- Hypothesis tests
- Extensions

2

Rationale

3

Panel Models

- What can we learn from datasets with many individuals but few time periods?
- Can we construct regression models based on panel datasets?
- What advantages do panel estimators have over estimates based on cross-sections alone?

4

Unobserved Heterogeneity

- Omitted variables bias
- Many individual characteristics are not observed
  - e.g. enthusiasm, willingness to take risks
- These vary across individuals – described as unobserved heterogeneity
- If these influence the variable of interest, and are correlated with observed variates, then the estimated effects of these variables will be biased

5

Applications of Panel Models

- Returns to Education
- Discrimination
- Informal caring
- Disability

6

Returns to education

- Cross-section estimates of returns to education
- Biased by failure to account for differences in ability?

7

Measurement of discrimination

- Gender/race discrimination in earnings may reflect unobserved characteristics of workers
- attitude to risk, unpleasant jobs etc.

8

One-way and two-way error components models

9

The Basic Data Structure

Wave 1

Individual 1

Wave T

Wave 1

Individual 2

Wave T

Wave 1

Individual N

Wave T

10

Formulate an hypothesis

11

Develop an error components model

Explanatory variables

Normally distributed error -

Constant across individuals

Composite error term

12

One-way or two-way error components?

Individual effect

Time Effect

Random error

13

Treatment of individual effects

Restrict to one-way model. Then two options for treatment of individual effects:

- Fixed effects – assume λ_i are constants
- Random effects – assume λ_i are drawn independently from some probability distribution

14

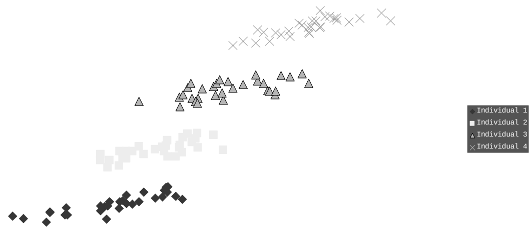
The Fixed Effects Model

Treat λ_i as a constant for each individual

λ now part of constant – but varies by individual

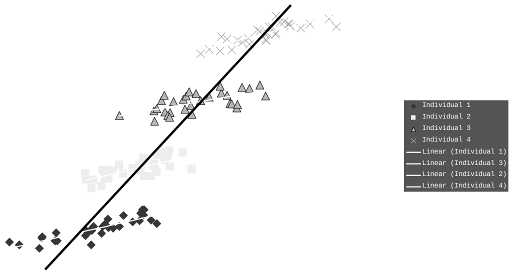
15

Graphically this looks like:



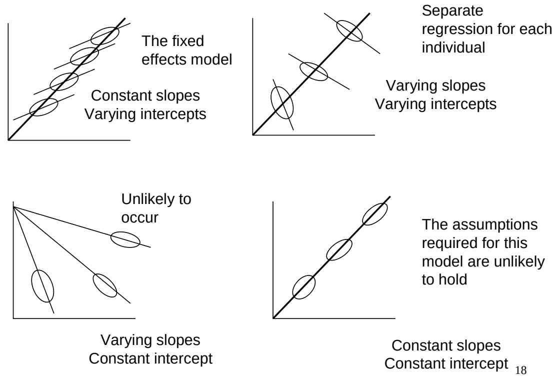
16

And the slope that will be estimated is BB rather than AA
Note that the slope of BB is the same for each individual
Only the constant varies



17

Possible Combinations of Slopes and Intercepts



18

Constructing the fixed-effects model - eliminating unobserved heterogeneity by taking first differences

Original equation

Lag one period and subtract

Constant and individual effects eliminated

Transformed equation

An Alternative to First-Differences:

Deviations from Individual Means

Applying least squares gives the first-difference estimator – it works when there are two time periods.
More general way of “sweeping out” fixed effects when there are more than two time periods - *take deviations from individual means*.
Let $x_{1i.}$ be the mean for variable x_1 for individual i, averaged across all time periods. Calculate means for each variable (including y) and then subtract the means gives:

The constant and individual effects are also eliminated by this transformation

Estimating the Fixed Effects Model

Take deviations from individual means and apply least squares – fixed effects, LSDV or “within” estimator

It is called the “within” estimator because it relies on variations within individuals rather than between individuals. Not surprisingly, there is another estimator that uses only information on individual means. This is known as the “between” estimator. The Random Effects model is a combination of the Fixed Effects (“within”) estimator and the “between” estimator.

Three ways to *estimate* β

overall

within

between

The overall estimator is a weighted average of the “within” and “between” estimators. It will only be *efficient* if these weights are correct.

The *random effects* estimator uses the **correct weights**.

22

The Random Effects Model

Original equation

Remember

λ_i now part of error term

This approach might be appropriate if observations are representative of a sample rather than the whole population. This seems appealing.

23

The Variance Structure in Random Effects

In random effects, we assume the λ_i are part of the composite error term ε_{it}. To construct an efficient estimator we have to evaluate the structure of the error and then apply an appropriate generalised least squares estimator to find an efficient estimator. The assumptions must hold if the estimator is to be efficient. These are:

This is a crucial assumption for the RE model. It is necessary for the consistency of the RE model, but not for FE. It can be tested with the Hausman test.

24

The Variance Structure in Random Effects

Derive the T by T matrix that describes the variance structure of the ε_{it} for individual *i*. Because the randomly drawn λ_i is present each period, there is a correlation between each pair of periods for this individual.

25

Random Effects (GLS Estimation)

The Random Effects estimator has the standard generalised least squares form summed over all individuals in the dataset i.e.

Where, given Ω from the previous slide, it can be shown that:

26

Fixed Effects (GLS Estimation)

The fixed effects estimator can also be written in GLS form which brings out its relationship to the RE estimator. It is given by:

Premultiplying a data matrix, X, by M has the effect of constructing a new matrix, X* say, comprised of deviations from individual means. (This is a more elegant way mathematically to carry out the operation we described previously) The FE estimator uses *M* as the weighting matrix rather than *Ω*.

27

Relationship between Random and Fixed Effects

The random effects estimator is a weighted combination of the "within" and "between" estimators. The "between" estimator is formed from:

28

Random or Fixed Effects?

For random effects:

•Random effects are efficient
•Why should we assume one set of unobservables fixed and the other random?
•Sample information more common than that from the entire population?
•Can deal with regressors that are fixed across individuals

Against random effects:

Likely to be correlation between the unobserved effects and the explanatory variables. These are assumed to be zero in the random effects model, but in many cases we might expect them to be non-zero. This implies **inconsistency** due to omitted-variables in the RE model. In this situation, fixed effects is inefficient, but still consistent.

29

Hypothesis Testing

- "Poolability" of data (Chow Test)
- Individual and fixed effects (Breusch-Pagan)
- Correlation between X_{it} and l_i (Hausman)

30

Test for Data Pooling

- Null (unconstrained) hypothesis – distinct regressions for each individual
- Alternative (constrained) – individuals have same coefficients, no error components (simple error)
- Appropriate test – F test (Chow Test)

31

Test for Individual Effects

- Breusch-Pagan Test
- Easy to compute – distributed as χ_2^2
- Tests of individual and time effects can be derived, each distributed as χ_1^2

32

The Hausman Test

Test of whether the Fixed Effects or Random Effects Model is appropriate

Specifically, test H_0: $E(\lambda_i|x_{it}) = 0$ for the one-way model

If there is no correlation between regressors and effects, then FE and RE are both consistent, but FE is inefficient.

If there is correlation, FE is consistent and RE is inconsistent.

Under the null hypothesis of no correlation, there should be no differences between the estimators.

33

The Hausman Test

A test for the independence of the λ_i and the x_{kit}.

The covariance of an efficient estimator with its difference from an inefficient estimator should be zero. Thus, under the null hypothesis we test:

If W is significant, we should not use the random effects estimator.

Can also test for the significance of the individual effects (Greene P562)

34

Extensions

- Unbalanced Panels
- Measurement Error
- Non-standard dependent variables
- Dynamic panels
- Multilevel modelling

35

Unbalanced Panels and Attrition

- Unbalanced panels are common and can be readily dealt with provided the reasons for absence are truly random.
- Attrition for systematic reasons is more problematic - leads to attrition bias.

36

Measurement Error

- Can have an adverse effect on panel models
- No longer obvious that panel estimator to be preferred to cross-section estimator
- Measurement error often leads to "attenuation" of signal to noise ratio in panels – biases coefficients towards zero

37

Non-normally distributed dependent variables in panel models

- Limited dependent variables - censored and truncated variables e.g. panel tobit model
- Discrete dependent variables – e.g. panel equivalents of probit, logit multinomial logit
- Count data – e.g. panel equivalents of poisson or negative binomial

38

Dynamic Panel Models

- Example - unemployment spell depends on
  - Observed regressor (e.g. *x* - education)
  - Unobserved effect (e.g. l – willingness to work)
  - Lagged effect (e.g. g - "scarring" effect of previous unemployment)

39

Multilevel Modelling

- Hierarchical levels
- Modelling performance in education
- Individual, class, school, local authority levels
- http://multilevel.ioe.ac.uk/

40

Multilevel Modelling

Equation has fixed and random component

Residuals at different levels

Individual j in school i attainment

41

Multilevel Modelling

Variance components model applied to JSP data
Explaining 11 Year Maths Score

Parameter	**Estimate (s.e.)**	**OLS Estimate (s.e.)**
Fixed:		
Constant	13.9	13.8
8-year score	0.65 (0.025)	0.65 (0.026)
Random:		
(between schools)	3.19 (1.0)	
(between students)	19.8 (1.1)	23.3 (1.2)
Intra-school correlation	0.14	

42

References

- Baltagi, B (2001) Econometric Analysis of Panel Data, 2nd edition, Wiley
- Hsiao, C. (1986) Analysis of Panel Data, Cambridge University Press
- Wooldridge, J (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press

43

Example from Greene's Econometrics Chapter 14
Open log, load data and check

```
log using panel.log
insheet using Panel.csv
edit
```

- Tell Stata which variables identify the individual and time period

```
iis i
tis t
```

44

Describe the dataset
xtdes

Now estimate the "overall" regression – ignores the panel properties

```
ge logc = log(c)
ge logq = log(q)
ge logf = log(pf)
regress logc logq logf
```

45

Calculate the “between” regression

```
egen mc = mean(logc), by(i)
egen mq = mean(logq), by(i)
egen mf = mean(logf), by(i)
egen mlf = mean(lf), by(i)
regress logc mq mf mlf
regress mc mq mf mlf lf
```

46

Calculate the “within” (fixed effects) regression

```
xtreg logc logq logf lf, i(i) fe
est store fixed
```

47

Equivalent to adding individual dummies (Least Squares Dummy Variables)

```
tabulate i, gen(i)
regress logc logq logf lf i2-i6
```

48

What do the dummy coefficients mean?

lincom _cons
lincom _cons + i2
lincom _cons + i3
lincom _cons + i4
lincom _cons + i5
lincom _cons + i6
regress logc logq logf lf i1-i6, noconst

49

Random effects

xtreg logc logq logf lf, i(i) re

50

Carry out Hausman test

hausman fixed

51