Panel Data Models Example

Ani Katchova

Panel Data Model Example

- We have panel data on wages and we need to account for variation over time (within) and across individuals (between). Log wage is the dependent variable and education, experience, and weeks worked are independent variables.
- Data from the Panel Study of Income and Dynamics (PSID).

Estimators:

- Pooled OLS
- Between
- Fixed effects (within)
- First differences
- Random effects

Within and between variation for panel data

Variable	Variation	Mean	St. Dev	Min	Max
Individual id	overall	298	171	1	595
	between		171	1	595
	within		0	298	298
Time t	overall	4	2	1	7
	between		0	4	4
	within		2	1	7
Log wage	overall	6.7	0.46	4.6	8.5
	between		0.39	5.3	7.8
	within		0.24	4.8	8.6
Education	overall	12.8	2.8	4	17
	between		2.8	4	17
	within		0	12.8	12.8
Experience	overall	19.9	11.0	1	51
	between		10.8	4	48
	within		2.0	17	23
Weeks worked	overall	47	5.1	5	52
	between		3.3	32	52
	within		3.9	12	64

- Time-invariant variables like individual id and education have positive between variation and zero within variation. Experience has more between variation (10.8) than within variation (2.0). The average experience of an individual is between 4 and 48 years across individuals, but varies by 7 years for each individual over time.

Comparing estimators for panel data models

Log wage	Pooled OLS regression	Between	Within or fixed effects	First differences	Random effects
Experience	0.04* (18)	0.03* (7)	0.11* (46)	0.11* (18)	0.08* (31)
Experience sq.	-0.0007* (-13)	-0.0006* (-5)	-0.0004* (-8)	-0.0005* (-4)	-0.0008* (-12)
Weeks worked	0.005* (4)	0.01* (3)	0.0008 (1.4)	-0.0003 (-0.5)	0.001 (1.3)
Education	0.007* (34)	0.07* (15)	dropped	dropped	0.11* (18)
Constant	4.90* (72)	4.68* (22)	4.60* (118)	no constant	3.83* (118)
R2	0.28			0.22	
R2-within		0.13	0.66		0.63
R2-between		0.32	0.03		0.17
R2-overall		0.27	0.05		0.18
Sigma u (α)			1.04		0.31
Sigma e			0.15		0.15
Rho			0.98		0.81
Theta (λ)					0.82

- Results show that higher values of experience and education are associated with higher values of log wages for all estimators while the weeks worked is significant for some estimators and not for others.
- OLS: across individuals and over time, an additional year of experience leads to 4% higher wages. Between: the average wages are 3% higher for individuals with one more year of average experience. Within: each additional year of experience above the average for an individual leads to 11% higher wages. First differences: each additional year increase in experience from one year to the next leads to 11% higher wages. Random: same as within.
- The Hausman test shows significant differences between the coefficients for the fixed effects and random effects model. Therefore, we need to use the fixed effects model.
- Some estimators do not provide coefficients for time-invariant regressors.
- Rho is the proportion of variation due to the individual specific term. We have a large proportion (98% or 81%) explained by the individual specific term and the rest due to idiosyncratic error.
- Lambda is 82%, so the RE estimates are much closer to the within estimates than to the OLS estimates.
- The R-squares show the between estimator can explain 32% of the between variation, and the fixed and random effects estimators can explain 66% and 63%, respectively of within variation.