Econometrics

Chapter 14: Panel Data Methods

In Choi

Sogang University

Fixed Effects Estimation

- When there are unobserved fixed effects, an alternative to first differences is fixed effects estimation.
- Consider the average over time of

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \mu_i + u_{it}.$$

The average of μ_i will be μ_i, so if you subtract the mean, μ_i will be differenced out just as when doing first differences. That is,

$$y_{it} - \bar{y}_{i.} = \beta_1 (x_{it1} - \bar{x}_{i.1}) + \ldots + \beta_k (x_{itk} - \bar{x}_{i.k}) + u_{it} - \bar{u}_{i.} \qquad (1)$$

- This method is also identical to including a separate intercept for every individual.

First Differences vs Fixed Effects

- First Differences and Fixed Effects will be exactly the same when $T = 2$. Why?
- For $T > 2$, the two methods are different.
- Fixed effects estimation is easily implemented for unbalanced panels (T different for each individual), not just for balanced panels.

Random Effects

- Start with the same basic model with a composite error

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \mu_i + u_{it}.$$

- Previously we assumed that μ_i was correlated with the x's, but what if it's not?
- OLS would be consistent in that case, but composite errors will be serially correlated.
- Need to transform the model and do GLS to solve the problem and make correct inferences.
- The GLS estimator is called the random effects estimator.

Random Effects

- End up with a sort of weighted average of OLS and Fixed Effects – use quasi-demeaned data.
- Let

$$\lambda = 1 - \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}}.$$

- The RE estimator is obtained by running regression on

$$y_{it} - \lambda \bar{y}_{i.} = \beta_1 (x_{it1} - \lambda \bar{x}_{i.1}) + \ldots + \beta_k (x_{itk} - \lambda \bar{x}_{i.k}) + v_{it} - \lambda \bar{v}_{i.},$$

  where $v_{it} = \mu_i + u_{it}$.
- If $\lambda = 1$, then this is just the fixed effects estimator.
- So, the bigger the variance of the unobserved effect, the closer it is to FE.

Random Effects

Example

Three Different Estimators of a Wage Equation (WAGEPAN.RAW; Wooldrige p.472)

Random Effects

Example

Dependent Variable: log (*wage*)			
Independent Variables	Pooled OLS	Random Effects	Fixed Effects
educ	.091 (.005)	.092 (.011)	
black	−.139 (.024)	−.139 (.048)	
hispan	.016 (.021)	.022 (.043)	
exper	.067 (.014)	.106 (.015)	
$exper^2$	−.0024 (.0008)	−.0047 (.0007)	−.0052 (.0007)
married	.108 (.016)	.064 (.017)	.047 (.018)
union	.182 (.017)	.106 (.018)	.080 (.019)

Random Effects

Example

(continued) Marriage premium falls to 4.7% when fixed effects estimator is used. This is consistent with the idea that marriage and unobserved ability are positively correlated. After taking into account individual ability, marriage still carries a premium. Why? Married people may be more stable in work places and work harder. Employers may be willing to pay for a premium for these possible reasons.

Fixed Effects or Random?

- If the individual effects and regressors are correlated, the FE estimator should be used. If not, the RE estimator is BLUE.
- Hausman's (1978) test can be used to find out whether the individual effects and regressors are correlated or not.
- The test statistic is based on FE-estimator - RE-estimator for the null hypothesis that the individual effects and regressors are nmot correlated.

Other Uses of Panel Methods

- It is possible to think of models where there is an unobserved fixed effect, even if we do not have true panel data.
- A well-known example is an unobserved family effect.
- See
  Ashenfelter and Krueger (1994), Estimates of the Economic Return to Schooling from a New Sample of Twins, *American Economic Review*, pp. 1157-1173.
- Can difference siblings.
- Can estimate family fixed effect model.

Other Uses of Panel Methods

Let
y_{1i}, y_{2i} : Logarithms of the wage rates of the first and second twin in the i-th couple
X_i : The set of variables that vary by family, but not across twins (age, race, measures of family background)
Z_{1i}, Z_{2i} : The set of variables that may vary across the twins (education levels, union status, job tenure, marital status)
μ_i : An unobservable components that vary by family.

Other Uses of Panel Methods

- The model is

$$\begin{aligned} y_{1i} &= \alpha' X_i + \beta' Z_{1i} + \mu_i + \varepsilon_{1i} \\ y_{2i} &= \alpha' X_i + \beta' Z_{2i} + \mu_i + \varepsilon_{2i}. \end{aligned} \tag{2}$$

  Here β denotes structural effects of the observables on earnings.

- Differencing the two equations in (2) gives

$$y_{1i} - y_{2i} = \beta'(Z_{1i} - Z_{2i}) + \varepsilon_{1i} - \varepsilon_{2i}.$$

  We may run OLS or IV on this equation. The IV estimation is considered due to the possible measurement errors in Z_{1i} and Z_{2i}.

Additional Issues

- Many of the theories we already know about both cross section and time series data can be applied to panel data.
- Can test and correct for serial correlation in the errors.
- Can test and correct for heteroskedasticity.
- Can estimate standard errors robust to both.