ECON 4230 Intermediate Econometric Theory

1 Panel Data

Panel data ...

- is the combination of cross-section and time series data.
- is also known as longitudinal or pooled data.
- increases your sample size and efficiency of your estimates.
- allows you to identify effects such as economies of scale and technological progress.

1.1 Panel Data Models

The basic population regression model is

$$Y_{it} = \alpha_i + \beta_2 X_{2,it} + \beta_3 X_{3,it} + u_{it}$$

where $i = 1, ..., N$ and $t = 1, ..., T$. The econometric models differ in how to treat α_i.

Pooled OLS

- $\alpha_i = \alpha$ for all $i = 1, ..., N$
- Draw figure

Fixed Effects (FE) Model

- α_i are treated as an unknown parameters (intercepts) that vary across i
- Draw figure
- Can be estimated with OLS and N intercept dummies
- FE model is also known as Least Squares Dummy Variable (LSDV) model
- $N + K$ parameters to estimate
- FE within-group estimator
- FE model cannot handle time-invariant regressors
- Restricted F test: H_o: Pooled OLS vs. H_A: FE Model

Random Effects (RE) Model

- Treat $\alpha_i = \alpha + \epsilon_i$ as a random variable with variance σ_ϵ^2
- RE model: $Y_{it} = \alpha + \beta_2 X_{2,it} + \beta_3 X_{3,it} + (\epsilon_i + u_{it})$
- RE model exhibits serial correlation
- GLS is the efficient estimator
- RE model is inappropriate (i.e., estimator is biased) if $corr(X_{it}, \epsilon_i) \neq 0$
- Hausman test: H_o: RE Model vs. H_A: FE Model

Application: U.S. Death Penalty