ANALYSIS OF PANEL DATA

Fixed-Effect and Random-Effect Models

datascience@berkeley

Fixed-Effect Model

A Digression: Differencing When There Are More Than Two Time Periods

Differencing with More than Two Time Periods

- Suppose we have N individuals and 3 time periods for each individual, totalling $3N$ observations. A general fixed effect model can be written as

$$y_{it} = \delta_1 + \delta_2 d2_t + \delta_3 d3_t + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + \epsilon_{it}$$

for $t = 1, 2, 3$

- The key assumption is that the error terms are uncorrelated with the explanatory variable in each time period:

$$\boxed{Cov(x_{itj}, \epsilon_{is})}$$

for all t, s, j

- This means that the explanatory variables are *strictly exogenous* after the unobserved effect a_i is eliminated.

- If an important time-varying variable is omitted from the model, then this assumption is violated.

Pause and Think (2 minutes):

Let's spend a couple of minutes to think about this assumption. It would help to write out the time index.

- The key assumption is that the error terms are uncorrelated with the explanatory variable in each time period:

$$Cov(x_{itj}, \epsilon_{is})$$

for all t, s, j

- This means that the explanatory variables are *strictly exogenous* after the unobserved effect a_i is eliminated.

- If a_i is correlated with x_{itj}, then x_{itj} will be correlated with the composite error, $v_{it} = a_i + \epsilon_{it}$.

- However, we can eliminate a_i by differencing adjacent periods.

- In the case where $T = 3$, we can substract time period one from time period two, time period two from time period three.

$$\triangle y_{it} = \delta_2 \triangle d2_t + \delta_3 \triangle d3_t + \beta_1 \triangle x_{it1} + \cdots + \triangle x_{itk} + \triangle \epsilon_{it}$$

for $t = 1, 2$

- If the equation satisfies the classical linear model assumptions, then pooled OLS gives unbiased estimators, and t and F statistics are valid for hypothesis testing. Asymptotic results can be used as well.

- As long as $\triangle \epsilon_{it}$ is uncorrelated with $\triangle x_{itj}$ for all j and $t = 2, 3$, then the OLS estimators are also consistent.

Berkeley SCHOOL OF INFORMATION