Fixed Effects (Within) Estimator
Econ 672

The Takeaway

The Fixed Effects Estimator can be a powerful identification strategy to satisfy the backdoor criterion (observed and unobserved confounders). We need panel data to develop this identification strategy, where we have the same unit (individual, firm, etc.) over time. We utilize the variation within the unit across time to estimate the impact of the treatment of interest on the outcome of interest. We can do this by demeaning the data or utilizing fixed effects estimator in Stata. When we demean the data, our time-invariant unobserved (and observed) heterogeneity drops out.

Pros:

An identification strategy that can control for observed and unobserved heterogeneity/confounders that do not vary over time.

Cons:

We cannot control for unobserved time-varying confounders (time-varying heterogeneity) and we cannot control for simultaneity or reverse causality.

Assumptions:
1) Strict exogeneity assumption: there are no time-varying confounders and the treatment not correlated with the error term and $E[\epsilon_{it}|D_{i1}, D_{i2}, \ldots, D_{iT}, u_i) = 0$
2) Regressors, treatment of interest, must vary over time or formally $rank(\sum_{t=1}^{T} E[\ddot{D}_{it}' \ddot{D}_{it}]) = K$

Testable Assumptions:

We cannot directly test the strict exogeneity assumption on unobserved time-varying confounders, but we control for time-varying observed confounders.

We can test to make sure that our treatment of interest and other regressors vary over time.

Estimation:
We can set up a model where we have unobserved time-invariant confounders. Where u_i is our unobserved time-invariant confounder for person i and our treatment of interest, D_{it}varies over time for person i at time i:

$$Y_{it} = \delta D_{it} + u_i + \epsilon_{it}$$

We take the mean of each variable across time across units:

$$\bar{Y}_{it} = \delta\bar{D} + u_i + \bar{\epsilon}_{it}$$

Demean the data:

$$(Y_{it} - \bar{Y}) = (\delta D_{it} - \delta\bar{D}) + (u_i - u_i) + (\epsilon_{it} - \bar{\epsilon})$$

Our fixed effects estimator utilizes demeaned data, which has dealt with our unobserved time-invariant confounder:

$$\ddot{Y}_{it} = \delta\ddot{D}_{it} + \ddot{\epsilon}_{it}$$

Assuming that our two assumption of Strict Exogeneity (no time-varying unobserved confounders) and our regressors must vary, we have an unbiased estimate of the treatment on the outcome of interest.

<u>Caveats of Fixed Effects</u>

These caveats include the following:

1) Fixed effects do not control for time-varying observed or unobserved confounders. We can include time-varying observed confounders in our model, but fixed effects cannot deal with time-varying unobserved confounders.

2) Fixed effects cannot control for simultaneity or reverse causality. We will need to utilize a different identification strategy in this scenario.

<u>Clustered Standard Errors</u>

One of the more important topics that we did not get to yesterday was what to do with standard errors with the Fixed Effects Estimator. We must cluster the standard errors by the panel units (e.g.: individuals, firms, etc.). When we cluster by the panel unit, this allows for εit to be correlated within the panel unit (e.g.: individuals, firms, etc.). If we do not cluster our standard errors will be biased. In addition, clustering will yield valid standard errors as long as the number of clusters is sufficiently large, where the rule of thumb is at least 30 clusters.