Fixed Effects Estimation

• When there is an unobserved fixed effect, an alternative to first differences is fixed effects estimation.

• Consider the average over time of

\[ y_{it} = \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it} \]

such that

\[ \frac{1}{T} \sum_{t=1}^{T} y_{it} = \beta_1 \frac{1}{T} \sum_{t=1}^{T} x_{it1} + \ldots + \beta_k \frac{1}{T} \sum_{t=1}^{T} x_{itk} + a_i + \frac{1}{T} \sum_{t=1}^{T} u_{it} \]

or using the notation

\[ \bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it} \]

and so on for the time averages we have

\[ \bar{y}_i = \beta_1 \bar{x}_{i,1} + \ldots + \beta_k \bar{x}_{i,k} + a_i + \bar{u}_i \]
and subtracting this equation from the first equation we end up with
\[ \begin{align*}
y_{it} - \bar{y}_i &= \beta_1 (x_{it,1} - \bar{x}_{i,1}) + \ldots + \beta_k (x_{it,k} - \bar{x}_{i,k}) + (u_{it} - \bar{u}_i) 
\end{align*} \]
such that the equation no longer has a fixed effect.

- The average of \( a_i \) will be \( a_i \), so if you subtract the mean, \( a_i \) will be differenced out just as when doing first differences

- Assume that \( k = 1 \) such that
\[ \begin{align*}
y_{it} - \bar{y}_i &= \beta_1 (x_{it,1} - \bar{x}_{i,1}) + (u_{it} - \bar{u}_i). \end{align*} \]
Then the Fixed Effects estimator of \( \beta_1 \) is simply the OLS estimator for the demeaned data:
\[ \begin{align*}
\hat{\beta}_{1,FE} &= \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it,1} - \bar{x}_{i,1}) (y_{it} - \bar{y}_i)}{\sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it,1} - \bar{x}_{i,1})^2}.
\end{align*} \]
• If we were to do this estimation by hand, we would need to be careful because we would think that the degrees of freedom are \( df = NT - k \), but in reality it is \( N(T - 1) - k \) because we used up df’s calculating means.

• Luckily, Stata (and most other packages) will do fixed effects estimation for you.

• This method is also identical to including a separate intercept for every individual. Example, assume that for individual \( i \) there are \( T_i \) time periods available. Then, considering the simple model with one regressor (\( k = 1 \)) we have

\[
\hat{\beta}_{1,FE} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T_i} (x_{it,1} - \bar{x}_{i,1}) (y_{it} - \bar{y}_i)}{\sum_{i=1}^{n} \sum_{t=1}^{T_i} (x_{it,1} - \bar{x}_{i,1})^2}
\]
and

$$\bar{y}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} y_{it}$$
First Differences versus Fixed Effects

• First Differences and Fixed Effects will be exactly the same when \( T = 2 \). To see this note that

\[
y_{i2} - \bar{y}_i = y_{i2} - \frac{1}{2} (y_{i2} + y_{i1}) = \frac{1}{2} (y_{i2} - y_{i1}) = \frac{1}{2} \Delta y_{i2}
\]

and

\[
y_{i1} - \bar{y}_i = y_{i1} - \frac{1}{2} (y_{i2} + y_{i1}) = -\frac{1}{2} (y_{i2} - y_{i1})
\]

so that after removing time averages we end up with two equations that are identical to the first difference formulation.

• For \( T > 2 \), the two methods are different

• Probably see fixed effects estimation more often than differences – probably more because it’s easier than that it’s better
• Theoretically, Fixed Effects estimation is better (in the sense of giving more precise parameter estimates, i.e. smaller standard errors) than First differences if the $u_{it}$ are serially uncorrelated.

• On the other hand First Differences are better if $\Delta u_{it}$ are serially uncorrelated. We have looked at tests and diagnostic statistics for checking the serial correlation of $\Delta u_{it}$. These tests can help to choose between FE and FD. Note that it is more difficult to test the serial correlation of $u_{it}$ because here we need to estimate each $a_{i}$.

• First Differences (estimated by OLS) are not appropriate when there is a lagged dependent variable included in the model. In other words, a model of the form

$$y_{it} = \phi y_{it-1} + \beta_{1} x_{it1} + a_{i} + u_{it}$$
should not be estimated by OLS in first difference form because $\Delta y_{it-1}$ is correlated with $\Delta u_{it}$. In this case, the fixed effects estimator is preferred. It will be biased when $T$ is small, but the bias disappears at the rate $T^{-1}$ in samples with more time periods. There are also simple bias corrections available for this case (Hahn and Kuersteiner, 2002, *Econometrica*).

- Fixed effects are easily implemented for unbalanced panels, not just balanced panels. Unbalanced panels are panels where individual cross-section units may have data for a different number of time periods.
A Wage Panel Example

- Vella and Verbeek (1998) analyze the National Longitudinal Survey (Youth Sample) from 1980 to 1987 for a sample of 545 men. This is a balanced panel data set with 8 years of data for each individual in the sample.

- Some variables in the data change over time: experience, marital status and union status are the three important ones.

- Other variables do not change: race and education are the key examples. If we use fixed effects or first differencing, we can not include race, education, or experience in the equation. This is because these variables are constant over time and thus get differenced out or can not be distinguished from the fixed effects.
However, we can include interactions of educ with year dummies for 1981 through 1987 to test whether the return to education was constant over this time period. We use log(wage) as the dependent variable, dummy variables for marital and union status a full set of year dummies and the interaction terms d81·educ, d82·educ,...,d87·educ.

The estimates on the interaction terms are consistent with an increase in the return to education over the period considered.
Random Effects

- Start with the same basic model with a composite error,

\[ y_{it} = \beta_0 + \beta_1 x_{i1t} + \ldots \beta_k x_{itk} + a_i + u_{it} \]

- Previously we’ve assumed that \( a_i \) was correlated with the \( x \)'s, but what if it’s not?

- OLS would be consistent in that case, but composite error will be serially correlated.

- To see this note that for

\[ v_{it} = a_i + u_{it} \]
where we assume that $E(u_{it}) = 0$, $Cov(u_{it}, u_{is}) = 0$ if $t \neq s$, $Cov(u_{it}, u_{js}) = 0$ if $i \neq j$.

- For $a_i$, $E(a_i) = 0$, $Cov(a_i, a_j) = 0$ if $i \neq j$ and $Cov(a_i, u_{jt})$ for all $i, j, t$ and most importantly, $Cov(a_i, x_{it,j}) = 0$.

- Then

$$E(v_{it}v_{is}) = \begin{cases} 
\sigma_a^2 & \text{for } t \neq s \\
\sigma_a^2 + \sigma_u^2 & \text{for } t = s 
\end{cases}$$

- Need to transform the model and do GLS (Generalized Least Squares) to solve the problem and make correct inferences

- Idea is to do quasi-differencing to remove the effect of serial correlation. The algebra to derive the required transformation is more advanced (matrix algebra) but the resulting estimator is readily available in Stata.
• Need to transform the model and do GLS (Generalized Least Squares) to solve the problem and make correct inferences.

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

\[
\lambda = 1 - \left( \frac{\sigma^2_u}{\sigma^2_u + T\sigma^2_a} \right)
\]

\[
y_{it} - \lambda \bar{y}_i = \beta_0 (1 - \lambda) + \beta_1 (x_{it,1} - \lambda \bar{x}_i,1) + \ldots
\]

\[
+ \beta_k (x_{it,k} - \lambda \bar{x}_i,k) + (v_{it} + \lambda \bar{v}_i)
\]

• If \( \lambda = 1 \), then this is just the fixed effects estimator

• If \( \lambda = 0 \), then this is just the OLS estimator
• So, the bigger the variance of the unobserved effect, the closer it is to FE

• The smaller the variance of the unobserved effect, the closer it is to OLS

• Stata will do Random Effects for us
Fixed Effects or Random?

- It is more common to think that we need fixed effects, since we think the problem is that something unobserved is correlated with the \( x \)'s

- One advantage of the random effects estimator is that it allows for the inclusion of regressors that are fixed over time. This includes variables such as education. However, since we believe that education is related to unobserved ability the random effects assumptions may not be that attractive after all.

- If the assumption of \( a_i \) being uncorrelated with \( x_{it} \) holds then the FE estimator is still valid, but less efficient (precise) than the RE estimator. In other words, if we are sure that random effects are appropriate it is better to use the RE estimator.
• If truly need random effects, the only problem is the standard errors. This means that we can run a pooled regression (rather than do the random effects transformation) and only correct for the standard error. The resulting estimator is less efficient than random effects, but will have correct standard errors.

• Can just adjust the standard errors for correlation within group.

• It is possible to test if a random effects specification is appropriate. This is done using a Hausman test. The idea is that fixed effects estimators are valid both when the $a_i$ are correlated with $x_{it}$ and when they are not correlated. The random effects estimator on the other hand is only valid if $a_i$ is not correlated with $x_{it}$. In the latter case it is more efficient than the fixed effects estimator.