17 Panel data. Fixed effects.

17.1 Time invariant heterogeneity

A panel is a set of observations on individuals, collected over time. An observation is the pair (y_{it}, x_{it}), where i denotes the individual and t denotes time. The standard panel data specification is that there is an individual-specific effect δ_i that is constant over time:

$$y_{it} = x_{it}\beta + \delta_i + \varepsilon_{it}$$

such that $\mathbb{E}[\varepsilon_{it}|x_{i1}, \ldots, x_{iT}, \delta_i] = 0$ (strict exogeneity). If δ_i is observed then we can estimate β by OLS of $y_{it} - \delta_i$ on x_{it}. If δ_i is unobserved then we require either the lack of correlation between δ_i and x_{it} or the availability of an instrument w_i which is correlated with x_{it} but not $\delta_i + \varepsilon_{it}$.

First differences

If neither of the above are available we can still consistently estimate β by considering the regression in first differences:

$$y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})'\beta + (\varepsilon_{it} - \varepsilon_{i,t-1}) \tag{17.1}$$

which is often written as the regression of Δy_{it} on Δx_{it}, where $\Delta y_{it} = y_{it} - y_{i,t-1}$ and $\Delta x_{it} = x_{it} - x_{i,t-1}$. Running OLS will not be optimal, since there is serial correlation in the error term. Thus the standard errors will be wrong and the estimator inefficient. To solve this we can use GLS.

17.2 Within-Group (or simply within) estimator

Let

$$y_{it} = x_{it}'\beta + \delta_i + \varepsilon_{it} \quad \text{where}$$

$$\underset{T\times 1}{y_i} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad \underset{T\times k}{x_i} = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{pmatrix}, \quad \underset{T\times 1}{\varepsilon_i} = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}$$

$$\underset{nT\times 1}{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{n1} \\ \vdots \\ y_{nT} \end{pmatrix}, \quad \underset{nT\times k}{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \underset{nT\times 1}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1T} \\ \vdots \\ \varepsilon_{n1} \\ \vdots \\ \varepsilon_{nT} \end{pmatrix}, \quad \underset{n\times 1}{\delta} = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix}$$

We assume strict exogeneity $\mathbb{E}[\varepsilon_i|x_i, \delta_i] = 0$ and $Var(\varepsilon_i|x_i, \delta_i) = \sigma^2 I_T$. We can rewrite equation (1) in terms of the $(T-1) \times T$ matrix D with $D_{tt} = -1, D_{t,t+1} = 1$, and all other entries zero:

$$Dy_i = Dx_i\beta + D\delta_i + D\varepsilon_i = Dx_i\beta + D\varepsilon_i$$

$$\underset{(T-1)\times T}{D} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \Rightarrow \underset{(T-1)\times 1}{Dy_i} = \begin{pmatrix} y_{i2} - y_{i1} \\ \vdots \\ y_{iT} - y_{i,T-1} \end{pmatrix} = \Delta y_i$$

$$\Rightarrow \underset{(T-1)\times k}{Dx_i} = \begin{pmatrix} x_{i2} - x_{i1} \\ \vdots \\ x_{iT} - x_{i,T-1} \end{pmatrix} = \Delta x_i$$

This setup can also be written in terms of the matrices $Y, X, \varepsilon, \delta$, by considering the block diagonal matrix of D. Our assumptions become $\mathbb{E}[\varepsilon|X,\delta] = 0$ and $Var(\varepsilon|X,\delta) = \sigma^2 I_{nT}$. Let ℓ be a column vector of ones of length T. Then we can write the model as:

$$Y = X\beta + C\delta + \varepsilon \quad \text{where} \quad C = \begin{pmatrix} \ell & 0 & \dots & 0 \\ 0 & \ell & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \ell \end{pmatrix}$$

$$\begin{pmatrix} D & 0 & \dots & 0 \\ 0 & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D \end{pmatrix} Y = \begin{pmatrix} D & 0 & \dots & 0 \\ 0 & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D \end{pmatrix} X\beta + \begin{pmatrix} D & 0 & \dots & 0 \\ 0 & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D \end{pmatrix} \varepsilon$$

Note:-

Definition: The Kronecker product of two matrices A and B, denoted as $A \otimes B$, is a matrix formed by taking each element of A and multiplying it by the entire matrix B.

Example: Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$. The Kronecker product $A \otimes B$ is given by:

$$A \otimes B = \begin{pmatrix} 1 \cdot B & 2 \cdot B \\ 3 \cdot B & 4 \cdot B \end{pmatrix} = \begin{pmatrix} 5 & 6 & 10 & 12 \\ 7 & 8 & 14 & 16 \\ 15 & 18 & 20 & 24 \\ 21 & 24 & 28 & 32 \end{pmatrix}$$

Properties

- $(A \otimes B)' = A' \otimes B'$
- $A \otimes (C + D) = A \otimes C + A \otimes D$

There are many other useful properties, but these are all we need for this lecture.

We can now write the first difference estimator more succinctly as:

$$(I_n \otimes D)Y = (I_n \otimes D)X\beta + (I_n \otimes D)\varepsilon$$

GM Assumptions

GM1 rank$(I_n \otimes D)$X = k

This is equivalent to rank(D)X = k, which is equivalent to rank(X) = k

GM2 $\mathbb{E}[\Delta\varepsilon|\Delta X] = 0$

$\mathbb{E}[(I_n \otimes D)\varepsilon|\Delta X] = (I_n \otimes D)\mathbb{E}[\varepsilon|\Delta X] = 0$

GM3 Homoskedasticity and no serial correlation in $\Delta\varepsilon$

Even if we assume that ε_{it} is homoskedastic and serially uncorrelated, $\Delta\varepsilon_{it}$ will not be. $Cov(\varepsilon_{it} - \varepsilon_{i,t-1}, \varepsilon_{i,t-1} - \varepsilon_{i,t-2}) = -Var(\varepsilon_{it}|x_i) = -\sigma^2 \neq 0$.

Since GM1 and GM2 hold, the FD below is unbiased and consistent.

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{n}(Dx_i)'(Dx_i)\right)^{-1}\left(\sum_{i=1}^{n}(Dx_i)'(Dy_i)\right)$$

However GM3 is violated so the OLS estimate is not BLUE. To get the efficient estimator we need to use GLS.

Definition 17.2.1: GLS

If instead of GM3 we have that $Var(\varepsilon|X) = \Omega$ where Ω is known, then we can transform the data by premultiplying by $\Omega^{-1/2}$ to get: $Y^* = \Omega^{-1/2}Y, X^* = \Omega^{-1/2}X, \varepsilon^* = \Omega^{-1/2}\varepsilon$. Then we can run OLS on $Y^* = X^*\beta + \varepsilon^*$ to get the efficient estimator $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$. Note that Ω can be scaled without any effect on the estimator.

We can see this results in efficient estimates by considering $Var(\varepsilon^*|X^*) = \Omega^{-1/2}Var(\varepsilon|X)\Omega^{-1/2} = \Omega^{-1/2}\Omega\Omega^{-1/2} = I_{nT}$. We now derive the GLS estimator for the FD model. We assume that the original regression satisfies $\mathbb{E}[\varepsilon|X] = 0$ and $Var(\varepsilon|X) = \sigma^2 I_{nT}$. Then,

$$\begin{aligned}
Var((I_n \otimes D)\varepsilon|X) &= (In \otimes D)Var(\varepsilon|X)(In \otimes D)' \\
&= \sigma^2 \begin{pmatrix} D & 0 & \dots & 0 \\ 0 & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D \end{pmatrix} \begin{pmatrix} D' & 0 & \dots & 0 \\ 0 & D' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D' \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} DD' & 0 & \dots & 0 \\ 0 & DD' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & DD' \end{pmatrix}
\end{aligned}$$

We thus have our covariance matrix to use for scaling,

$$\Omega = \sigma^2 \begin{pmatrix} DD' & 0 & \dots & 0 \\ 0 & DD' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & DD' \end{pmatrix} \Rightarrow \Omega^{-1/2} = \frac{1}{\sigma} \begin{pmatrix} (DD')^{-1/2} & 0 & \dots & 0 \\ 0 & (DD')^{-1/2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (DD')^{-1/2} \end{pmatrix}$$

Thus,

$$
\begin{aligned}
\hat{\beta}_{GLS-FD} &= \left(((I_n \otimes D)X)'\Omega^{-1}(I_n \otimes D)X\right)^{-1}((I_n \otimes D)X)'\Omega^{-1}(I_n \otimes D)Y \\
&= \left(X'(I_n \otimes D)'(I_n \otimes (DD')^{-1})(I_n \otimes D)X\right)^{-1} X'(I_n \otimes D)'(I_n \otimes (DD')^{-1})(I_n \otimes D)Y \\
&= \left(X'(I_n \otimes D')(I_n \otimes (DD')^{-1})(I_n \otimes D)X\right)^{-1} X'(I_n \otimes D')(I_n \otimes (DD')^{-1})(I_n \otimes D)Y \\
&= \left(X'(I_n \otimes D'(DD')^{-1}D)X\right)^{-1} X'(I_n \otimes D'(DD')^{-1}D)Y
\end{aligned}
$$

Note that $D'(DD')^{-1}D$ is a projection matrix onto the space spanned by columns of D' (and thus is symmetric and idempotent). We now consider the column vector of 1s ℓ. Since the FD of ℓ is zero (i.e. always $1-1$), we have that $D'\ell = 0$. Thus a projection onto the space spanned by D' is equivalent to projection onto the space orthogonal to ℓ.

$$
\Rightarrow D'(DD')^{-1}D = I_T - \ell(\ell'\ell)^{-1}\ell' = \frac{1}{T}\ell\ell'
$$

Then,

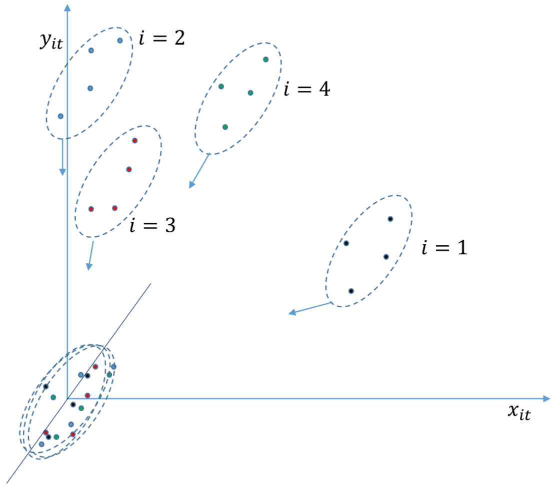
$$
\begin{aligned}
\hat{\beta}_{GLS-FD} &= \left(X'(I_n \otimes (I_T - \frac{1}{T}\ell\ell'))X\right)^{-1} X'(I_n \otimes (I_T - \frac{1}{T}\ell\ell'))Y \\
&= \left(\sum_{i=1}^{n} x_i'(I_T - \frac{1}{T}\ell\ell')x_i\right)^{-1}\left(\sum_{i=1}^{n} x_i'(I_T - \frac{1}{T}\ell\ell')y_i\right) \\
&= \left(\sum_{i=1}^{n} x_i'(x_i - \ell\frac{\ell' x_i}{T})\right)^{-1}\left(\sum_{i=1}^{n} x_i'(y_i - \ell\frac{\ell' y_i}{T})\right)
\end{aligned}
$$

$$
x_i - \ell\frac{\ell' x_i}{T} = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{pmatrix} - \frac{1}{T}\begin{pmatrix} \ell' x_i \\ \vdots \\ \ell' x_i \end{pmatrix} = \begin{pmatrix} x_{i1} - \frac{1}{T}\sum_{t=1}^{T} x_{it} \\ \vdots \\ x_{iT} - \frac{1}{T}\sum_{t=1}^{T} x_{it} \end{pmatrix} = \underbrace{\begin{pmatrix} x_{i1} - \bar{x}_i \\ \vdots \\ x_{iT} - \bar{x}_i \end{pmatrix}}_{T\times k}
$$

$$
\begin{aligned}
\hat{\beta}_{GLS-FD} &= \left(\sum_{i=1}^{n} x_i'(x_i - \bar{x}_i)\right)^{-1}\left(\sum_{i=1}^{n} x_i'(y_i - \bar{y}_i)\right) \\
&= \left(\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)\right)
\end{aligned}
$$

This estimator is known as the within or within-group estimator because it exploits only the variation of x_{it} and y_{it} within the groups of observations corresponding to the same individual but different time periods. It ignores the variability between groups by subtracting individual specific time averages from the data. Note that the errors are not uncorrelated, indeed every error term appears in every other equation (via the demeaning).

A geometric interpretation of the within transformation is given here. The individual effects are the 'intercepts' of the individual specific scatters. Here they are negatively correlated with x_it so fitting that regression line to the untransformed data results in negative bias. The within transformation shifts the data so that the individual specific scatters have zero intercepts, and thus the regression line is unbiased.



END OF LECTURE