1 Panel Data

- In many empirical investigations it is common to have repeated observations on the same unit and this then gives us variation across time and across individuals.
- In this way it is common to write the model as

$$y_{it} = \beta_0 + x_{it}\beta + c_i + u_{it}$$

  where c is an individual specific effect.
- In the treatment of these models it is common to assume

$$E[u_t|x_t, c] = 0...t = 1..T$$

- One implication of this condition is

$$E[u_t|x_t] = 0.$$

- If we were to assume $E[x_t'c] = 0$ then one could also apply OLS to the above model.
- However, if the individual effect is correlated with the $x's$ then OLS would not be biased.

- In proceeding to estimate the parameters of this model it is important that we clarify our treatment of the individual components.
- An important distinction is whether they are treated as random or fixed.
- We will discuss this more in detail below.

- Another important issue is the exogeneity assumptions regarding the $x's.$
- It is very restrictive to assume the $x's$ are non random as this excludes potential feedback from y_{it} to x_{is} for $s > t.$
- With an unobserved effect the clearest form of strict exogeneity is

$$E[y_{it}|x_{i1}, x_{i2}, ...x_{iT,} c_i] = E[y_{it}|x_{it}, c_i] = x_{it}\beta + c_i$$

  where the second equality is a functional form assumption.

- When this above assumption holds (ie. the first equality) we say that the $x's$ are strictly exogenous conditional on the unobserved effect.
- Generally the strict exogeneity assumption is written in terms of the idiosyncractic errors as

$$E(u_{it}|x_{i1}, x_{i2}, ...x_{iT,} c_i) = 0$$

- This assumes that the explanatory variables in each time period are uncorrelated with the idiosyncractic errors in each time period.
- That is,
$$E(x'_{is}u_{it}) = 0, \ \ s, t = 1...T$$
- This is much stronger than contemporaneous correlation.
- However, note that it does not impose any restrictions on the relationship between x and c.

1.1 Estimation by Pooled OLS

- Write the model as

$$y_{it} = x_{it}\beta + v_{it}$$

- where

$$v_{it} = c_i + u_{it}$$

- We know that OLS is consistent if $E(x'_{it}v_{it}) = 0$ which essentially implies that

$$E(x'_{it}u_{it}) = E(x'_{it}c_i) = 0$$

- Note that even the estimates will be consistent it is necessary to adjust the standard errors to account for the correlation which is induced since c_i is included in each error for the same individual.

1.2 Random Effects Estimation

- It is useful, from an efficiency perspective, to account for the presence of c_i in the calculation of the standard errors and this is essentially what random effects estimation does.
- However, in doing so it imposes additional assumptions above what is required for consistency than pooled OLS.
- Assumption RE.1: $a) E(u_{it}|x_i, c_i) = 0, t = 1..T;\ b) E(c_i|x_i) = E(c_i) = 0$ noting that $x_i = (x_{i1}, ..x_{iT})$.

- We can see that strict exogeneity is implied by the above. Under the above assumption it follows that

$$E(v_{it}|x_i) = 0$$

- Write the model for all T time periods as

$$y_i = X_i\beta + v_i$$

and v_i can be written as $v_i = c_i j_T + u_i$ where j_T is the Tx1 vector of ones

- Define the unconditional variance matrix of v_i as

$$\Omega = E(v_i v_i')$$

which is a TxT matrix that is positive definite.

- For consistency of GLS we require the usual rank condition for GLS
- Assumption RE.2: rank $E(X\prime_i \Omega^{-1} X_i) = K$

- Now, a feasible GLS estimator using a consistent estimate of Ω would be consistent but this would not exploit the individual specific component of the error term.
- A standard random effects estimator imposes structure which exploits this feature.

- The first assumption is that the unconditional variance of the idiosyncratic term is constant across t. That is,

$$E(u_{it}^2) = \sigma_u^2,\ t = 1..T$$

- The second assumption is that the idiosyncratic error terms are uncorrelated.

$$E(u_{it}u_{is}) = 0,\ \text{all } t \neq s$$

- Under these two assumptions we can derive the variances and covariances of the elements of v_i.
- Under RE.1a $E(c_i u_{it}) = 0$ and thus

$$E(v_{it}^2) = E(c_i^2) + 2E(u_{it}c_i) + E(u_{it}^2) = \sigma_c^2 + \sigma_u^2$$

  Also, for all $t \neq s$

$$E(v_{it}v_{is}) = E[(c_i + u_{it})(c_i + u_{is})] = \sigma_c^2.$$

- Therefore under the above assumptions

$$\Omega = E(v_i v_i') = \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & . & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & . & . \\ . & . & . & . \\ \sigma_c^2 & . & & \sigma_c^2 + \sigma_u^2 \end{bmatrix}$$

- Since $j_T j_T'$ is the TxT matrix with unity in every element, we can write the above as

$$\Omega = \sigma_u^2 I_T + \sigma_c^2 j_T j_T'.$$

- The Ω above has a random effects structure and rather than depending on T(T+1)/2 unrestricted variances and covariances it only depends on the two unknown parameters σ_c^2 and σ_u^2.

- We can summarize the assumptions as
- Assumption RE.3: a) $E(u_i u_i' | x_i, c_i) = \sigma_u^2 I_T$. $b) E(c_i^2 | x_i) = \sigma_c^2$.
- Under RE.3.a. $E[u_{it}^2 | x_i, c_i] = \sigma_u^2$, $t = 1..T$ and $E[u_{it} u_{is} | x_i, c_i] = 0, t \neq s, t, s = 1...T$. RE3.a. assumes the conditional variances are constant and the conditional covariances are zero.
- Assumption RE.3.b is the same as the homoskedasticity assumption on the unobserved effect.

- To implement a feasible GLS procedure define

$$\sigma_v^2 = \sigma_c^2 + \sigma_u^2$$

  and assume, for now, that we have consistent estimates of σ_c^2 and σ_u^2.

- Thus we can form

$$\widehat{\Omega} = \widehat{\sigma}_u^2 I_T + \widehat{\sigma}_c^2 j_T j_T' \quad (1)$$

  which is a TxT matrix that we assume to be positive definite.

- The FGLS estimator that uses 1 is known as the random effects estimator.

$$\widehat{\beta}_{RE} = \left(\sum_{i=1}^{N} X_i' \widehat{\Omega}^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' \widehat{\Omega}^{-1} y_i \right). \tag{2}$$

- To implement this RE estimator we need estimates of the variances which we assumed to have above.

- It is easiest to first obtain $\widehat{\sigma}_v^2 = \widehat{\sigma}_u^2 + \widehat{\sigma}_c^2$.
- Under assumption RE3a. $\sigma_v^2 = T^{-1}\sum_{t=1}^{T} E(v_{it}^2)$ for all i.
- Thus averaging v_{it}^2 across all i and t gives a consistent estimate of σ_v^2.
- However, this requires an initial estimate of β and one can use the pooled OLS estimate $\widehat{\widehat{\beta}}$.

- Let $\widehat{v}_{it}$ denote the pooled OLS residuals.
- A consistent estimator of σ_v^2 is given by

$$\widehat{\sigma}_v^2 = \frac{1}{(NT-K)} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\widehat{v}}_{it}^2.$$

- To find a consistent estimator of σ_c^2 recall that $\sigma_c^2 = E(v_{it}v_{is})$, all $t \neq s$.
- Therefore for each i, there are are $T(T-1)/2$ non redundant error products that can be used to estimate σ_c^2.

- If we sum all these combinations and take the expectation, we get, for each i

$$E\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} v_{it}v_{is}\right) = \sum_{t=1}^{T-1}\sum_{s=t+1}^{T} E\left(v_{it}v_{is}\right) = \sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \sigma_c^2 = \sigma_c^2 \sum_{t=1}^{T-1}(T-t)$$

$$= \sigma_c^2((T-1)+(T-2)+......2+1) = \sigma_c^2 T(T-1)/2 \quad (3)$$

where we have used the fact that the sum of the first T-1 positive integers is T(T-1)/2.

- As usual a consistent estimator is obtained by replacing the expectation with an average (across i) and replacing v_{it} with its pooled OLS residual.
- We also make a DOF adjustment as a small sample correction:

$$\widehat{\sigma}_c^2 = \frac{1}{[NT(T-1)/2 - K]} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \widehat{\widehat{v}}_{it} \widehat{\widehat{v}}_{is}$$

  is a consistent estimator of σ_c^2 under Assumptions RE.1-RE.3.
- Given that we have $\widehat{\sigma}_c^2$ and $\widehat{\sigma}_v^2$ we can form $\widehat{\sigma}_u^2 = \widehat{\sigma}_v^2 - \widehat{\sigma}_c^2$.

1.3 Fixed Effects Estimation

- Recall that the model under investigation for T time periods has the form

$$y_{it} = x_{it}\beta + c_i + u_{it} \; t = 1..T. \tag{4}$$

- The random effects estimator essentially puts the c_i in the error term under the assumption that it is orthogonal to the x_{it} and then accounts for the implied correlation in the error term from doing so.

- However, in many economic studies the inclusion of the c_i is precisely to capture that there is some unobservable which is potentially correlated with the $x's.$
- The fixed effects estimator is designed to account precisely for this possibility. The T equations can be written as

$$y_i = X_i\beta + c_i j_T + u_i \tag{5}$$

where the j_T is the $Tx1$ vector of ones.

- This equation represents a single draw from the cross section.

- The first fixed effects assumption is strict exogeneity of the explanatory variables conditional on c_i :
- Assumption FE.1.

$$E(u_{it}|x_i, c_i) = 0, \ t = 1, 2..T.$$

- One can immediately see that this type of assumption regarding the individual effect is far more appealing than that of the random effects procedures.
- However, this increased level of flexibility is associated with a decrease in the number of parameters that are identified.
- That is, as we allow for any type of relationship between the fixed effects and the $x's$ this means the coefficients on any time invariant x is not identified.
- Also note that this assumption retains the strict exogeneity of the x.

- The idea for estimating β under the fixed effects assumption is to transform the equations in order to eliminate the c_i.
- When we have at least two observations on the same i we can eliminate c_i through an appropriate "fixed effects transformation" or "within transformation".
- To perform the FE transformation we first average over 4 to get

$$\overline{y}_i = \overline{x}_i\beta + c_i + \overline{u}_i \tag{6}$$

where $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_i$, etc.

- Subtracting this from the original model we get

$$y_{it} - \overline{y}_i = (x_{it} - \overline{x}_i)\beta + (u_{it} - \overline{u}_i) \tag{7}$$

or

$$\widetilde{y}_{it} = \widetilde{x}_{it}\beta + \widetilde{u}_{it} \;\; t = 1..T. \tag{8}$$

where $\widetilde{y}_{it} = y_{it} - \overline{y}_i$ etc.

- The time demeaning of the original equation has removed the individual effect c_i.

- The OLS estimation of 8, which represents FE estimation, provides consistent estimates of β provided

$$E(\widetilde{x}_{it}'\widetilde{u}_{it}) = 0 \tag{9}$$

  noting that in general this condition does not hold if we relax strict exogeneity.

- We can also write the model 8 for all time periods

$$\widetilde{y}_i = \widetilde{X}_i\beta + \widetilde{u}_i \tag{10}$$

  where $\widetilde{y}_i$ is Tx1, $\widetilde{X}_i$ is a Txk and $\widetilde{u}_i$ is Tx1.

- This set of equations can be obtained by premultiplying 5 by a time demeaning matrix . Define $Q_T = I_T - j_T(j_T'j_T)^{-1}j_T'$ which is a TxT symmetric, idempotent matrix with rank T-1.

- Note that $Q_Tj_T = 0;\ Q_Ty_i = \widetilde{y}_i;\ Q_TX_i = \widetilde{X}_i;\ Q_Tu_i = \widetilde{u}_i$ and so premultiplying 5 by Q_T gives the demeaned equations is 10.

- For the FE estimator to be well behaved we require

- Assumption FE.2: rank $\left(\sum_{t=1}^{T} E(\widetilde{x}_{it}'\widetilde{x}_{it})\right) = rank\left[E(\widetilde{X}_i'\widetilde{X}_i)\right] = K.$

- Thus if the x_{it} contains an element that does not vary over t for any i then the corresponding element in $\widetilde{x}_{it}$ is zero and the above assumption does not hold. Thus the coefficients on time invariant variables are not identified.

- The fixed effects estimator can be expressed as

$$\begin{aligned}\widehat{\beta}_{FE} &= \left(\sum_{i=1}^{N} \widetilde{X}_i' \widetilde{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \widetilde{X}_i' \widetilde{y}_i\right) \qquad (11) \\ &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{x}_{it}' \widetilde{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{x}_{it}' \widetilde{y}_{it}\right).\end{aligned}$$

- This is also called the within estimator because it uses the time variation with within each cross section.
- The between estimator, which uses only variation between the cross section observations , is the OLS estimator applied to the time averaged equation 6.
- This estimator is not consistent under assumption FE.1 because $E(\overline{x}_i' c_i)$ is not necessarily zero.

- So far we have treated the c_i as unobserved random variables. However, often we treat the c_i as parameters to be estimated.
- If assumption FE.2 is changed to its finite sample version, rank$(\tilde{X}'\tilde{X}) = K$ then the models satisfies the gauss markov assumptions conditional on X.

- To estimate each c_i along with the parameters β one possibility is to define N dummy variables, one for each cross sectional unit, such that

$$
\begin{aligned}
dn_i &= 1 \text{ if } n = i \\
dn_i &= 0 \text{ otherwise}
\end{aligned}
$$

  and then regress y_{it} on $x_{it},\ d1_i, ..dn_i.$

- The $\widehat{c}_1$ is the coefficient on $d1_i$ etc.

- This regression also corresponds to the FE estimator described above and this is why the FE estimator is also known as the dummy variable estimator.

- There is an important distinction between the $\widehat{c}_i$ and the $\widehat{\beta}_{FE}$.
- We know that the $\widehat{\beta}_{FE}$ is consistent with T fixed as $N \to \infty$.
- This is not the case with $\widehat{c}_i$ as each additional cross sectional unit (i.e. N increasing) means that an additional parameter that has to be estimated and information does not accumulate on any c_i as N increases.
- Thus each $\widehat{c}_i$ is an unbiased estimator of c_i when the c_i are treated as parameters.
- This is a practical example of an estimator which is unbiased but not consistent.