0.1. Panel Data.

- Suppose we have a panel of data for groups (e.g. people, countries or regions) $i = 1, 2, ..., N$ over time periods $t = 1, 2, ..., T$ on a dependent variable y_{it} and a $kx1$ vector of independent variables x_{it} and we are interested in measuring the effect of x_{it} on y_{it}. say

$$y_{it} = \alpha_i + \beta_i' x_{it} + \varepsilon_{it} \tag{0.1}$$

- where β_i is a $kx1$ vector and $E(\varepsilon_{it}) = 0$; $E(\varepsilon_{it}^2) = \sigma_i^2$; $E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}$; $E(\varepsilon_{it}\varepsilon_{jt-s}) = 0$ for $s \neq 0$. Notice that here k does not include the intercept, whereas above it did.

The panel data estimators for the linear model are all standard, either the application of OLS or GLS.

- There are 3 literatures on this type of problem, distinguished by the relative magnitudes of N and T and the assumptions that are made about parameter and variance homogeneity.

1. The large T small N literature. This uses time-series asymptotics, T going to infinity N fixed. The standard model is the Zellner Seemingly Unrelated Regression Estimator, SURE, which estimates the full model above by GLS allowing for the between group covariances $E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}$. Notice that the between group covariance matrix involves estimating N(N+1)/2 elements, so grows rapidly with N.

2. The large N small T literature. This arises typically with large surveys like the BHPS where the number of time periods is small (5 is quite large) but there may be many thousand cross-section observations. T is not large enough to estimate a model for each group so strong homogeneity assumptions tend to be imposed on the slope parameters $\beta_i = \beta$ and also often on the intercept parameters, $\alpha_i = \alpha$. Between group covariances, σ_{ij}, are asumed to be zero. The asymptotic properties of the estimators are established by letting N go to infinity, T fixed. These are usually non-linear models.

3. Large N large T literature (sometimes known as panel time-series), where T is large enough to estimate an equation for each group, but N is too large to allow for a freely estimated between group covariance matrix. The asymptotics involves letting both N and T go to infinity in some way.

1. Fixed Effects

The most widely used model in the panel literature is the Fixed Effect (FE) model:

$$y_{it} = \alpha_i + \beta' x_{it} + u_{it} \tag{1.1}$$

$E(u_{it}) = 0; E(u_{it}^2) = \sigma^2$ all i; $E(u_{it}u_{jt-s}) = 0$ for $s \neq 0$ and $i \neq j$. This restricts the slope coefficients and the variances to be the same across groups, while letting the intercepts differ, and treats all between group covariances as zero. This model is known by a large number of different names, because it was developed independently in many areas. These include:

the "Least Squares Dummy Variable" model (because it can be implemented by running a least squares regression including a dummy (0,1) variable for each group;

the "Within Estimator" since it just uses the within group variation, see below;

the (one way) "Fixed Effects" estimator, in contrast to the two way Fixed Effects and Random Effects estimators discussed below;

the analysis of covariance estimator; and various other names.

Notice that we cannot estimate α_i consistently ($N \to \infty, T$ fixed), since the number of parameters grows with the sample size N. However we can estimate β consistently.

The total variation in y_{it} can be decomposed into the within group variation and the between group variation:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \overline{y})^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \overline{y}_i)^2 + T\sum_{i=1}^{N}(\overline{y}_i - \overline{y})^2$$

where

$$\overline{y} = \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}/NT; \quad \overline{y}_i = \sum_{t=1}^{T} y_{it}/T;$$

the FE "Within Regression" just uses the within group variation, since the group specific intercepts can be removed by taking deviations from the group mean, allowing (1.1) to be written:

$$(y_{it} - \overline{y}_i) = \beta'(x_{it} - \overline{x}_i) + u_{it} \tag{1.2}$$

the "Between regression" is the cross-section regression using the group means:

$$\overline{y}_i = \alpha + \beta' \, \overline{x}_i + u_i.$$

If the intercepts are all regarded as identical, then one just gets standard OLS on all the data:

$$y_{it} = \alpha + \beta^{'} x_{it} + u_{it} \tag{1.3}$$

or

$$(y_{it} - \overline{y}) = \beta'(x_{it} - \overline{x}) + u_{it}. \tag{1.4}$$

This gives equal weight to the within group and the between group variation.

1.0.1. Parameterisation and two way models.

The parameters of dummy variable models like these can be written in a number of ways and this often gives cause for confusion. In (1.1) We estimate N parameters α_i, a separate intercept for each group. We could also express this as $\alpha_i = \alpha + \mu_i$, where $\sum_{i=1}^{N} \mu_i = 0$, i.e. estimating α and the $N - 1$ independent μ_i :

$$y_{it} = \alpha + \mu_i + \beta' x_{it} + u_{it}$$

This is equivalent to (1.2). Then, defining $\eta_i^x = \overline{x}_i - \overline{x}$, $\eta_i^y = \overline{y}_i - \overline{y}$,we can also express (1.2)

$$(y_{it} - \overline{y}_i + \overline{y} - \overline{y}) = \beta'(x_{it} - \overline{x}_i + \overline{x} - \overline{x}) + u_{it} \tag{1.5}$$

$$(y_{it} - \eta_i^y - \overline{y}) = \beta'(x_{it} - \eta_i^x - \overline{x}) + u_{it}. \tag{1.6}$$

The Two way fixed effect model allows for a separate intercept for every group and every time period:

$$y_{it} = \alpha_i + \alpha_t + \beta' x_{it} + u_{it}.$$

Notice that we cannot estimate N+T free intercepts (there would be exact multicolinearity, the dummy variable trap), some restriction is required to identify the parameters and a common one is to express the model as.

$$y_{it} = \alpha + \mu_i + \mu_t + \beta' x_{it} + u_{it}$$

subject to $\sum_{i=1}^{N} \mu_i = 0$, $\sum_{t=1}^{T} \mu_t = 0$. This can be estimated by taking deviations from the year means $\overline{y}_t$ and $\overline{x}_t$ and well as the group means, or

$$(y_{it} - \eta_i^y - \eta_t^y - \overline{y}) = \beta'(x_{it} - \eta_i^x - \eta_t^x - \overline{x}) + u_{it}. \tag{1.7}$$

.

An alternative restriction is to take one group-year as the base and express all the others as deviations from that.

1.0.2. Random Effect Models

The one way fixed effect model involves estimating N separate α_i and if N is large, in the thousands, this involves a lot of parameters and a large loss in efficiency. The alternative is the "Random Effects" model which treats the μ_i not as fixed parameters to be estimated, but as random variables, $E(\mu_i) = 0, E(\mu_i^2) = \sigma_\mu^2$. It is assumed that randomness implies that the μ_i are distributed independently of u_{it} and (the strong assumption) independently of x_{it}.

With these assumptions we only have to estimate 2 parameters α and σ_μ^2 not the N α_i. The model is then:

$$y_{it} = \alpha + \beta' x_{it} + (u_{it} + \mu_i)$$

where the parentheses indicate the new error term, $v_{it} = (u_{it} + \mu_i)$. $E(v_{it}) = 0; E(v_{it}^2) = \sigma^2 + \sigma_\mu^2$; $E(v_{it}v_{it-s}) = \sigma_\mu^2$, $s \neq 0$; $E(v_{it}v_{jt-i}) = 0$, $i \neq j$. Thus this error structure introduces a very specific form of serial correlation. Estimation is by Generalised Least Squares, equivalent to OLS on the transformed equation:

$$(y_{it} - \theta \overline{y}_i) = \beta'(x_{it} - \theta \overline{x}_i) + u_{it},$$

where;

$$\theta = 1 - \frac{\sigma}{\sqrt{T\sigma_\mu^2 + \sigma^2}}.$$

Notice that the Fixed Effect estimator corresponds to the case where $\theta = 1$ or $T\sigma_\mu^2$ is infinite. The Random Effect Estimator lies between the FE and OLS estimates. Feasible GLS requires an estimate of θ, but there are a number of possible consistent estimates from the FE or OLS first stage regressions. You can also have random time effects.

1.1. Testing

If the unrestricted model is, the general model with no between group covariances

$$y_{it} = \alpha_i + \beta_i' x_{it} + \varepsilon_{it} \tag{1.8}$$

and the restricted model is the fixed effect model

$$y_{it} = \alpha_i + \beta' x_{it} + u_{it}. \tag{1.9}$$

This appears to be merely involve testing equality of the slope coefficients, i.e. the $k(N-1)$ restrictions $\beta_i = \beta$, $i = 1, 2, ..., N$. The standard F test (Chow test) for this is:

$$\frac{(\sum\sum \widehat{u}_{it}^2 - \sum\sum \widehat{\varepsilon}_{it}^2)/N(k-1)}{\sum\sum \widehat{\varepsilon}_{it}^2/(NT - N(k+1))} \sim F[k(N-1), (NT - N(k+1))].$$

The difficulty is that this test will only be correct if the variances are the same across groups: $\sigma_i^2 = \sigma$ for all i.

One alternative is to use Likelihood Ratio Tests which can be calculated from the same two sets of least squares regressions. If both coefficients and variances differ, the maximised log likelihood is, apart from a constant, the sum of the log likelihoods for the individual equations:

$$L_1 = -\frac{T}{2}\sum_{i=1}^{N} \ln \widehat{\sigma}_i^2; \quad \widehat{\sigma}_i^2 = \sum_{t=1}^{T} \widehat{\varepsilon}_{it}^2.$$

If the coefficients differ, but the variances are the same, the maximised log likelihood is:

$$L_2 = -\frac{NT}{2} \ln \widehat{\sigma}^2; \quad \widehat{\sigma}^2 = \sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{\varepsilon}_{it}^2.$$

If both the slope coefficients and the variances are the same, the maximised log likelihood is:

$$L_3 = -\frac{NT}{2} \ln \widetilde{\sigma}^2; \quad \widetilde{\sigma}^2 = \sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{u}_{it}^2.$$

There is a fourth case, equal coefficients and different variances, discussed above.

The test for equality of variances is then just $2(L_1 - L_2) \sim \chi^2(N-1)$. The test for equality of both coefficients and variances is just $2(L_1 - L_3) \sim \chi^2(N-1)(k+1)$. The LR equivalent of the F test above (equality of coefficients conditional on equality of variances) is $2(L_2 - L_3) \sim \chi^2(k[N-1])$. Exactly the same sort of procedure can be used for testing equality of intercepts. If N was small we could start from the more general SURE model, which allowed for between group covariances.

The Likelihood ratio approach does not work with the Random Effects model, since it is a GLS rather than ML estimator. The usual approach is to test between OLS and RE, by using a standard LM test for heteroskedasticity, since the variances will differ between groups if the RE model is appropriate. Of course, you may get heteroskedasticity for other reasons than random effects.

To test between RE and FE a Hausman test is used. Call the FE estimator $\widehat{\beta}^F$ with estimated Variance Covariance matrix $V(\widehat{\beta}^F)$ and the RE estimator $\widehat{\beta}^R$ with $V(\widehat{\beta}^R)$. If the RE model is correct, $\widehat{\beta}^R$ is consistent and efficient so $V(\widehat{\beta}^F) > V(\widehat{\beta}^R)$. The variance of the difference is $V(\widehat{\beta}^F - \widehat{\beta}^R) = V(\widehat{\beta}^F) - V(\widehat{\beta}^R)$. If the RE model is wrong (the effects are not random but correlated with the x_{it}) then the RE estimates are inconsistent, but the FE estimates are still consistent. The Hausman tests uses as a test statistic:

$$(\widehat{\beta}^F - \widehat{\beta}^R)'[V(\widehat{\beta}^F) - V(\widehat{\beta}^R)]^{-1}(\widehat{\beta}^F - \widehat{\beta}^R) \sim \chi^2(k)$$

If this is large (the difference between the estimates is significant) you reject the null hypothesis that the RE model is appropriate against the alternative that the FE model is appropriate.

1.1.1.

2. Dynamics.

- Consider a dynamic version of the fixed effect model

$$y_{it} = \alpha_i + \beta^{'} x_{it} + \lambda y_{i,t-1} + u_{it} \tag{2.1}$$

- the usual estimator is inconsistent ($N \to \infty$, T fixed), because of the usual problem of the downward bias of the lagged dependent variable because of dependence on initial conditions, though the bias declines with T.

- There are various instrumental variable estimators which are consistent, which remove the α_i by differencing rather than by taking deviations from the group means.

- If you difference, you get

$$\Delta y_{it} = \beta^{'} \Delta x_{it} + \lambda \Delta y_{i,t-1} + \Delta u_{it} \tag{2.2}$$

- but $\Delta u_{it} = u_{it} - u_{i,t-1}$ is clearly correlated with $\Delta y_{i,t-1} = y_{it} - y_{i,t-1}$ since $u_{i,t-1}$ determines $y_{i,t-1}$.

- However, you can use $y_{i,t-2}$ and earlier as instruments.

- Suppose the coefficients differ:

$$y_{it} = \alpha_i + \beta_i^{'} x_{it} + \lambda_i y_{i,t-1} + u_{it} \tag{2.3}$$

- and this is ignored, then the equation is:

$$y_{it} = \alpha_i + \beta^{'} x_{it} + \lambda y_{i,t-1} + [(\beta_i - \beta)^{'} x_{it} + (\lambda_i - \lambda) y_{i,t-1} + u_{it}] \tag{2.4}$$

- where $v_{it} = [(\beta_i - \beta)^{'} x_{it} + (\lambda_i - \lambda) y_{i,t-1} + u_{it}]$ is the new error term.
- This error term is going to be serially correlated and correlated with the lagged dependent variable, so the estimates will be inconsistent even for large T.
- This heterogeneity bias, biases $\widehat{\lambda}$ upwards (i.e. in the opposite direction to the initial condition bias) and $\widehat{\beta}$ downwards towards zero.
- If x_{it} is I(1) serial correlation coefficient of unity and $\lambda_i = 0$, $P\lim(\widehat{\beta}) = 0$, $P\lim(\widehat{\lambda}) = 1$, $T \to \infty$. The estimates are wrong.
- The bias on the long-run coefficients $\beta/(1-\lambda)$ is smaller because the two biases cancel out to some extent.

2.0.2. Alternatives to Pooling

- If T is large enough to estimate an equation for each group, rather than imposing homogeneity, which can have adverse consequences (particularly in dynamic models), it may be better to use a weighted average of the individual coefficients of the form

$$\widetilde{\beta} = \sum W_i \widehat{\beta}_i [\sum W_i]^{-1}$$

- There are a number of such estimators available, differing in the choice of weights.

- One the Swamy Random Coefficients Models, weights inversely to the adjusted variances of the $\widehat{\beta}_i$.

- Note that the FE estimator is also of this form where

$$W_i = [\ \widetilde{\sigma}^2 (X_i' X_i)^{-1}]^{-1}.$$