ECON 21410 Spring 2019

A Review of Econometrics with Panel Data

Ari Boyarsky (aboyarsky@uchicago.edu)

Preface

The following are a condensed set of notes on the econometric analysis of panel data. The notes loosely follow the exposition given by Bruce Hansen's Econometrics. All mistakes are my own.

Contents

1 What is Panel Data?

Loosely speaking panel data are data where the observations for a particular "individual" exist over multiple periods. It is also known as longitudinal data.

In Economics this type of data is often used because it allows us to control for unobserved time-invariant endogeniety without the use of an instrument. It also allows us to capture heterogeneity within individuals in the data.

Example 1.1 (Firm Data over Years)

Firm ID	Year	Revenue
1	1990	...
1	2000	
2	1990	
2	2000	
3	1990	
3	2000	...

Some notation: we let $i = 1, \ldots, N$, S_i denote the time periods attributed to each individual i, $t = 1, \ldots, T_i$ for individual i. Notice that, $X_i \in \mathbb{R}^{T_i \times K}$ and $X = (X_1^T, \ldots, X_N^T)$.

2 Pooled Regression

I begin with an exposition of the pooled regression. This should also serve as a reminder of some of the properties of OLS.

The pooled regression takes the form,

$$y_{i,t} = x_{i,t}^T \beta + \epsilon_{i,t} \tag{1}$$

Along with the assumption that, $\mathbb{E}[X_{it} e_{it}] = 0$. Then $\hat{\beta}_{Pooled}$ is given by,

$$\hat{\beta}_{Pooled} = \left(\sum^N \sum_{t \in S_i} x_{it} x_{it} \right)^{-1} \left(\sum^N \sum_{t \in S_i} x_{it} y_{it} \right) = (X^T X)^{-1} (X^T Y) \tag{2}$$

as usual. We have that $\epsilon_i = y_i - X_i \hat{\beta}$. Ideally we want to make the following assumption,

Assumption 2.1 (Strict Mean Independence) $\mathbb{E}[\epsilon_{it} | X_i] = 0$

We differentiate this from $\mathbb{E}[\epsilon_{it} | X_{it}] = 0$ which allows for dependence over time (i.e. lagged variables). Notice that we also make the relevant rank assumption throughout.

Now we claim that the pooled estimate is unbiased. You should remind yourself what this means.

Theorem 2.2 *$\hat{\beta}_{Pooled}$ is an unbiased estimator.*

Proof: Just as in typical OLS we have,

$$\begin{aligned}
\hat{\beta}_{Pooled} &= \left(\sum^{N} X_i X_i\right)^{-1} \left(\sum^{N} X_i (X_i \beta + \epsilon_i)\right) \\
&= \beta + \left(\sum^{N} X_i X_i\right)^{-1} \left(\sum^{N} X_i \epsilon_i\right) \\
\implies \mathbb{E}[\hat{\beta}_{Pooled}|X] &= \beta + 0 \qquad\qquad \text{Using } \mathbb{E}[\epsilon_i|X] = 0
\end{aligned}$$

■

If we have homoskedasticity and no serial correlation then the variance estimator is the usual homoskedastic estimator for OLS. Otherwise, we want to use a clustered standard error estimate,

$$\hat{V}(\hat{\beta}_{Pooled}) = (X^T X)^{-1} (\sum^{N} X_i \epsilon_i \epsilon_i X_i)(X^T X)^{-1} \tag{3}$$

3 Random Effects

In the pooled regression, consider a decomposition of the error term such that $\epsilon_{it} = u_i + e_{it}$. The u_i is an individual error term and the e_{it} are typical iid errors. The random effects estimator uses this decomposition and makes the following assumption.

Assumption 3.1 (Random Effects Specification)

$$\mathbb{E}[e_{it}|X_i] = 0,\ \mathbb{E}[e_{it}^2|X_i] = \sigma_e^2,\ \mathbb{E}[e_{ij}e_{it}|X_i] = 0,\ \mathbb{E}[u_i|X_i] = 0,\ \mathbb{E}[u_i^2|X_i] = \sigma_u^2,\ \mathbb{E}[u_i e_{it}|X_i] = 0\ \ \forall j \neq t$$

In general this assumption implies that $\mathbb{E}[\epsilon_i|X_i] = 0$ and that $\mathbb{E}[\epsilon_i \epsilon_i|X_i] = \mathbb{1}_{T_i}\mathbb{1}_{T_i}\sigma_u^2 + I_{T_i}\sigma\epsilon^2$

The random effects regression is then,

$$y_{it} = x_{it}\beta + u_i + e_{it} \tag{4}$$

To give some brief intuition we make the assumption that the individual term is uncorrelated with the other covariates. Thus, we can allow for time-invariant variables to have an effect in our model. As we will see in the next section, this can lead to OVB if the individual terms are correlated through some unknown mechanism. Fixed effects lets us relax this assumption but comes at a cost of time-invariant variables in our data.

Where the error term follows Assumption 3.1. Estimation of the random effects model is usually done using GLS. Recall,

$$\hat{\beta}_{GLS} = \left(\sum_{i=1}^{N} X_i^T \Omega_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i^T \Omega_i^{-1} y_i\right) \tag{5}$$

We can compute the Ω matrix using the error structure defined in Assumption 3.1. In particular, $\Omega_i = I_{T_i} + \mathbb{1}_{T_i}\mathbb{1}_{T_i}\sigma_u^2/\sigma_e^2$. However, in practice we do not know these quantities. So, we can define a sample analogue. First define the projection matrix $P_{T_i} = \mathbb{1}_{T_i}(\mathbb{1}_{T_i}^T\mathbb{1}_{T_i})^{-1}\mathbb{1}_{T_i}^T$ and the residual maker matrix $M_{T_i} = I_{T_i} - P_{T_i}$.

Notice that these are both orthogonal projection matrices and so idempotent ($P^2 = P$) and orthogonal ($PP^T = I \implies P^{-1} = P^T$).

$$\begin{aligned} \Omega_i &= I_{T_i} + \mathbb{1}_{T_i}\mathbb{1}_{T_i}\sigma_u^2/\sigma_e^2 = I_{T_i} + \frac{T_i\sigma_u^2}{\sigma_e^2}P_{T_i} \quad \text{Using } T_i = \mathbb{1}_{T_i}^T\mathbb{1}_{T_i} \\ &= M_{T_i} + \rho_{T_i}^{-2}P_{T_i} \end{aligned} \tag{6}$$

Where the last equality follows from,

$$\rho_{T_i} = \frac{\sigma_e}{\sqrt{\sigma_e^2 + T_i\sigma_u^2}} \tag{7}$$

Such that, $\rho_{T_i}^{-2} = 1 + \frac{T_i\sigma_u^2}{\sigma_e^2}$. So that,

$$M_{T_i} + \rho_{T_i}^{-2}P_{T_i} = I_{T_i} - P_{T_i} + (1 + \frac{T_i\sigma_u^2}{\sigma_e^2})P_{T_i} = I_{T_i} + \frac{T_i\sigma_u^2}{\sigma_e^2}P_{T_i}$$

Of course, ρ_{T_i} is unknown so we can replace it with a sample analogue,

$$\hat{\rho}_{T_i} = \frac{\hat{\sigma}_e}{\sqrt{\hat{\sigma}_e^2 + T_i\hat{\sigma}_u^2}} \tag{8}$$

Notice that while this is in itself seems trivial the fact that we may do so is a key point in statistics that relies on Assumption 3.1 (specifically, mean independence, homoskedasticity, and uncorrelated errors). Using these assumptions and assuming a fixed T to simplify the representation, we can estimate the sample variances as,

$$\hat{\sigma}_\epsilon^2 = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{n}\hat{e}_{it} \tag{9}$$

$$\hat{\sigma}_u^2 = \frac{1}{NT - N}\sum_{t=1}^{T}\sum_{i=1}^{N}(\hat{e_{it}} - \frac{\sum_{t=1}^{T}\hat{e}_{it}}{T}) \tag{10}$$

$$\hat{\sigma}_e^2 = \sigma_\epsilon^2 - \sigma_u^2 \tag{11}$$

Where we can calculate $\hat{e}_{it} = y_{it} - \beta_0 - x_{it}\beta - u_i\gamma$. Then we simply calculate,

$$\begin{aligned} \tilde{y}_i &= y_i - (1-\hat{\rho})1_{T_i}\bar{y}_i \\ \tilde{X}_i &= X_i - (1-\hat{\rho})1_{T_i}\bar{x}_i \end{aligned} \tag{12}$$

And finally, we simply compute $(\tilde{X}^T\tilde{X})^{-1}(\tilde{X}^T\tilde{y})$.

4 Fixed Effects

The main difference between fixed and random effects is that in the fixed effects model we allow u_i, the fixed effect, to be correlated with x_{it}. Indeed, in many economic applications this is often the case. Essentially, the random effects estimator allows for a random intercept but the same slope because it needs to be mean independent from the error term. If this does not hold we produce biased estimates. Now consider the fixed effects model,

$$y_{it} = x_{it}\beta + u_i + \epsilon_{it} \tag{13}$$

In order to get identification we need,

Assumption 4.1 (Strict Exogenity)

$$\mathbb{E}[x_{is}\epsilon_{it}] = 0 \ \ \forall s = 1, \ldots, T$$

Notice that $\mathbb{E}[\epsilon_{it}|X_i] = 0$ is a stronger assumption.

4.1 The Within Transformation

In this section we will see that in order to allow the fixed effects to be correlated with the Xs and consistently estimate β we can use a demeaning operation known as a within transformation. Define the mean for a variable to be,

$$\bar{y_i} = \frac{1}{T_i}\sum_{t=1}^{T_i} y_{it} \tag{14}$$

And then the within transformation takes the form,

$$\dot{y}_{it} = y_{it} - \bar{y}_i$$

It is also useful to consider this with matrix algebra,

$$\begin{aligned}
\dot{y}_i &= y_i - \mathbb{1}\bar{y}_i \\
&= y_i - \mathbb{1}_{T_i}(\mathbb{1}_{T_i}^T \mathbb{1}_{T_i})^{-1}\mathbb{1}_{T_i}^T y_i \\
&= M_i y_i
\end{aligned}$$

Now, consider the formulation,

$$\bar{y}_i = \bar{x}_i\beta + u_i + \bar{\epsilon}_i \tag{15}$$

Notice, we have demeaned every variable except u_i which is already at the individual level. Now if we subtract 13 by this we remove the individual terms,

$$\dot{y}_i = \dot{x}_i\beta + \dot{\epsilon}_i \tag{16}$$

Now we will give some intuition for what is going on here. Essentially, we have removed all time-invariant trends from our problem. The idea being that we can capture all time-invariant endogeneity by demeaning over the periods an individual is observed. This also implies that any time-invariant regressors will also be implied. This is actually implied by the model because these regressors can be correlated with our fixed-effects term and so will be captured in it's estimates.[1]

[1]The Kennedy panel notes uploaded to Github provides further intuition and a further comparison with random effects.

4.2 The Fixed Effects Estimator

Now that we have built the within transformation we can consider the actual fixed effects estimator given by,

$$\begin{aligned}\hat{\beta}_{FE} &= \left(\sum^N \dot{X}_i \dot{X}_i\right)^{-1}\left(\sum^N \dot{X}_i \dot{y}_i\right)\\ &= \left(\sum^N X_i M_{T_i} X_i\right)^{-1}\left(\sum^N X_i M_{T_i} y_i\right)\end{aligned}$$

Where the application of the residual maker matrix follows form the exposition in the last section. Notice that this now assumes $\dot{X}^T\dot{X}$ is full rank.

Theorem 4.2 *$\hat{\beta}_{FE}$ is unbiased.*

Proof: Notice that just as in typical OLS we can write,

$$\hat{\beta}_{FE} - \beta = \left(\sum^N X_i M_{T_i} X_i\right)^{-1}\left(\sum^N X_i M_{T_i} \epsilon_i\right)$$

Then, taking conditional expectations yields,

$$\mathbb{E}[\hat{\beta}_{FE}|X] = \left(\sum^N X_i M_{T_i} X_i\right)^{-1}\left(\sum^N X_i M_{T_i} \mathbb{E}[\epsilon_i|X]\right) + \beta = \beta$$

Under the mean independence assumption. ■

We can also calculate the variance of the fixed effects estimator. Let,

$$\Sigma_i = \mathbb{E}[\epsilon_i \epsilon_i | X_i]$$

Then the variance of the estimator is given by,

$$V(\hat{\beta}_{FE}) = (\dot{X}^T\dot{X})^{-1}(\sum^N \dot{X}_i \Sigma_i \dot{X}_i)(\dot{X}^T\dot{X})^{-1} \tag{17}$$

Theorem 4.3 (Asymptotic Theory for FE) *Equation 17 gives the asymptotic variance of the FE estimator.*

Proof: Again write,

$$\hat{\beta}_{FE} - \beta = \left(\sum^N X_i M_{T_i} X_i\right)^{-1}\left(\sum^N X_i M_{T_i} \epsilon_i\right)$$

Then,

$$\sqrt{N}(\hat{\beta}_{FE} - \beta) = \left(\frac{1}{n}\sum^N \dot{X}_i \dot{X}_i\right)^{-1}\frac{1}{\sqrt{N}}\left(\sum^N \dot{X}_i \dot{\epsilon}_i\right)$$

Notice, now that by applying the LLN to the numerator and denominator we get that,

$$\left(\frac{1}{n}\sum^{N}\dot{X}_i\dot{X}_i\right)^{-1} \underset{N}{\rightarrow} \mathbb{E}[\dot{X}_i^T\dot{X}_i]^{-1}$$

And,

$$\left(\frac{1}{\sqrt{n}}\sum^{N}\dot{X}_i\dot{\epsilon}_i\right) \underset{N}{\rightarrow} \mathbb{E}[\dot{X}_i^T\dot{\epsilon}_i]^{-1}$$

So that the central limit theorem yields,

$$\left(\frac{1}{\sqrt{n}}\sum^{N}\dot{X}_i\dot{\epsilon}_i\right) \rightarrow N(0, \mathbb{E}[\dot{X}_i^T\dot{\epsilon}_i\dot{\epsilon}_i\dot{X}_i])$$

Finally, applying Slutsky's theorem (If you haven't seen this no problem! This is just what allows us to do the multiplication during the distributional convergence) gives us that,

$$\sqrt{N}(\hat{\beta}_{FE} - \beta) \underset{d}{\rightarrow} \mathbb{E}[\dot{X}_i^T\dot{X}_i]^{-1}N(0, \mathbb{E}[\dot{X}_i^T\dot{\epsilon}_i\dot{\epsilon}_i\dot{X}_i])$$

So that the asymptotic variance is,

$$avar(\hat{\beta}_{FE}) = \mathbb{E}[\dot{X}_i^T\dot{X}_i]^{-1}\mathbb{E}[\dot{X}_i^T\dot{\epsilon}_i\dot{\epsilon}_i\dot{X}_i]\mathbb{E}[\dot{X}_i^T\dot{X}_i]^{-1}$$

■

If the errors are homoskedastic ($\mathbb{E}[\epsilon_{it}^2|X_i] = \sigma_\epsilon^2)$) and serially uncorrelated ($\mathbb{E}[\epsilon_{ij}\epsilon_{it}|X_i] = 0$) then we have,

$$V(\hat{\beta}_{FE}) = \sigma_\epsilon^2(\dot{X}^T\dot{X})^{-1} \tag{18}$$

One important detail to notice is that we do suffer a cost for using fixed effects over pooled regression and random effects. In particular fixed effects admits a higher variance since we have reduced the variation in the regressors by effectively demeaning. That is, we suffer a loss of statistical efficiency. However, we gain <u>alot</u>! In particular, we are now robust to individual effects in our data.