USC, Fall 2016, Economics 513

Lecture 9: Panel data: Random Effects

Panel data

If we have multiple observations on the members of a population we have panel data. Examples

- In a year a (random) sample of firms is selected and these firms are followed over time. Such data are also called longitudinal data. A famous panel survey for households is the Panel Study of Income Dynamics (PSID).
- We collect readily available data on US states for a number of years. This gives longitudinal data on states.
- We collect data on pupils in a random sample of classes.
- We visit a twins festival and obtain data by interviewing the (identical) twins present.

Advantages and disadvantages of panel data

- An important advantage of panel data is that we can more easily deal with omitted (and often unobservable) variables.
- In a firm panel we can control for management quality; in a state panel for the political culture in a state; in a class panel for teacher quality; in an (identical) twins panel for the genetic endowment.
- This is important if we e.g. are interested in estimating the effect of an increase in the sentencing laws in a state on the crime rate in that state. In a cross section of states high crime states may already have high minimum penalties which biases the estimated effect. Using panel data we can control for that.
- If we are interested in the change in household income we get more accurate estimates from a panel survey.
- A disadvantage is that a panel study may suffer from dropout, i.e. panel attrition, or from panel conditioning in the responses to survey questions.

We must distinguish between panel data and repeated crosss ections in which e.g. in subsequent years we draw random samples from the same population. An example is the Current Population Survey (CPS) that is a repeated cross-section of US households.

Repeated cross sections do not have panel attrition and can be used instead of panel data in some cases.

Linear regression models with unobserved effects

For each member of the population indexed by i we have T observations. If we have the same number of observations we have a balanced panel. If the number of observations is different for different i then we have an unbalanced panel.

It is important to know why a panel is unbalanced. For instance, if in a household panel a household is more likely to drop out if its income falls, then we may get a biased estimate of the change in household income.

In a balanced panel the observations are y_{it}, x'_{it} for $t = 1, \ldots, T, i = 1, \ldots n$. We consider the unobserved effects model

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}$$

The variables α_i capture omitted variables that do not vary with t but do vary with i. Therefore the composite random error is $\eta_{it} = \alpha_i + \varepsilon_{it}$.

The method we use to estimate the regression coefficients depends on the assumption we make on the random error η_{it}. Regarding α_i we can assume that

- $\mathrm{E}(\alpha_i | x_{i1}, \ldots, x_{iT}) = 0$
- α_i and $x_{i1}, \ldots, x_{iT}$ are dependent.

They are dependent if e.g. $x_{i1}, \ldots, x_{iT}$ are chosen by i taking α_i into account. Examples: Fertilizer choice by farmer with α_i soil quality or education choice by twins with α_i innate ability.

Initially we assume that $\mathrm{E}(\alpha_i | x_{i1}, \ldots, x_{iT}) = 0$. This is usually called the random effects assumption.

Random effects

We organize the data by i, i.e.

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} \qquad X_i = \begin{pmatrix} x'_{i1} \\ \vdots \\ x'_{iT} \end{pmatrix} \qquad \eta_i = \begin{pmatrix} \eta_{i1} \\ \vdots \\ \eta_{iT} \end{pmatrix}$$

Consider the random error vector η_i. Then $\eta_{it}\eta_{is}$ are correlated (they share α_i) but η_{it} and η_{js} are not if $i \neq j$. Without further assumptions we have

$$\mathrm{E}(\eta_i \eta'_i | X) = \Sigma(X_i) \qquad \mathrm{E}(\eta_i \eta'_j | X) = 0$$

with $\Sigma(X_i)$ a $T \times T$ matrix and X the $nT \times K$ matrix of the independent variables.

Because the RE assumption implies that the error is mean independent of the independent variables, the main issue is what we assume about $\Sigma(X_i)$. The individual effect random error $\eta_{it} = \alpha_i + \varepsilon_{it}$ implies(with some additional assumptions) that this matrix has a particular structure that we can exploit (see the RE estimator below). All estimators that we consider are consistent (no asymptotic bias), so that we are only concerned about the correct standard errors of the estimators. For all estimators we will give expressions of the so called robust variance matrix that does not require assumptions on $\Sigma(X_i)$.

The pooled OLS estimator

The OLS estimator is

$$\hat{\beta} = (X'X)^{-1}X'y = \left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x'_{it}\right)^{-1}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}y_{it}\right) =$$

$$\left(\sum_{i=1}^{n} X'_i X_i\right)^{-1}\left(\sum_{i=1}^{n} X'_i y_i\right)$$

If we substitute the model we get from the first expression on the right hand side

$$\hat{\beta} = \beta + \left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x'_{it}\right)^{-1}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}\eta_{it}\right) =$$

$$\beta + \left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T} x_{it}x'_{it}\right)\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T} x'_{it}\eta_{it}\right)\right)$$

We use this expression to show consistency of OLS if we assume

$$\mathrm{E}(\eta_{it}|x_{it}) = 0$$

or even

$$\mathrm{E}(\eta_{it}x_{it}) = 0$$

We apply the LLN for $n \to \infty$, because under either assumption

$$\mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} x'_{it}\eta_{it}\right] = 0$$

For the asymptotic distribution we consider the second expression on the right hand side after substitution of the model

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} X'_i X_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X'_i \eta_i\right)$$

By the usual argument we find that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$$

where V can be estimated by

$$\hat{V} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i' X_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} X_i' \hat{\eta}_i \hat{\eta}_i' X_i\right) \left(\frac{1}{n}\sum_{i=1}^{n} X_i' X_i\right)^{-1}$$

with

$$\hat{\eta}_i = y_i - X_i \hat{\beta}$$

Note that we did not make any assumption on the (conditional) covariance of η_{it} and η_{is}, i.e. $\Sigma(X_i)$ is unrestricted. The OLS estimator for panel data is called the pooled OLS estimator. It is consistent under the assumption made, but the usual formula for the (asymptotic) variance of the OLS estimator does not apply.

The random effects estimator

Given the interpretation of η_{it} The precise nature of the correlation between η_{it} and η_{is} depends on the assumption on the correlation of ε_{it} and ε_{is}.

We assume

$$\mathrm{E}(\varepsilon_{it}\alpha_i|X_i) = 0 \qquad \mathrm{E}(\varepsilon_{it}\varepsilon_{is}|X_i) = 0 \; t \neq s \qquad \mathrm{E}(\alpha_i^2|X_i) = \sigma_\alpha^2 \qquad \mathrm{E}(\varepsilon_{it}^2|X_i) = \sigma_\varepsilon^2$$

This is a set of (conditional) uncorrelatedness and (conditional) variance (i.e. homoskedasticity) assumptions.

With $\eta_{it} = \alpha_i + \varepsilon_{it}$ this implies that

$$\mathrm{E}(\eta_{it}^2|X_i) = \mathrm{E}(\alpha_i^2|X_i) + \mathrm{E}(\varepsilon_{it}^2|X_i) + 2\mathrm{E}(\alpha_i\varepsilon_{it}|X_i) = \sigma_\alpha^2 + \sigma_\varepsilon^2$$

and for $s \neq t$

$$\mathrm{E}(\eta_{it}\eta_{is}|X_i) = \mathrm{E}(\alpha_i^2|X_i) + \mathrm{E}(\alpha_i\varepsilon_{is}|X_i) + \mathrm{E}(\alpha_i\varepsilon_{it}|X_i) + \mathrm{E}(\varepsilon_{it}\varepsilon_{is}|X_i) = \sigma_\alpha^2$$

Therefore (ι_T is a T vector of 1's)

$$\mathrm{Var}(\eta_i|X_i) = \sigma_\alpha^2 \iota_T \iota_T' + \sigma_\varepsilon^2 I_T \equiv \Omega$$

Remember that in the CLR model that has a diagonal variance matrix of the random errors of the form $\sigma^2 I$, the OLS estimator was optimal, i.e. BLU (Gauss-Markov theorem). This suggests that we can improve on pooled OLS if we 'transform' the data such that the variance matrix of the transformed random error is diagonal and has constant variance and apply OLS to the transformed data. If we write

$$y_i = X_i\beta + \eta_i$$

then we should find a $T \times T$ matrix A such that the variance matrix of $A\eta_i$ is diagonal. This is a general idea that applies if the random errors are correlated and have varying variances. The estimator based on the (linearly) transformed data is the Generalized Least Squares (GLS) estimator, i.e. we use OLS in

$$Ay_i = AX_i\beta + A\eta_i$$

with A such that

$$\mathrm{E}(A\eta_i\eta_i A'|X_i) = A\Omega A' = \sigma_\varepsilon^2 I_T$$

With an obvious notation we take $A = \Omega^{-1/2}$ and in this case

$$\Omega^{-1/2} = I_T - \frac{\lambda}{T}\iota_T\iota_T'$$

with

$$\lambda = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}}$$

Check that $\Omega^{-1/2}$ is symmetric and that $\Omega^{-1/2}\Omega\Omega^{-1/2} = \sigma_\varepsilon^2 I_T$.

By this transformation the dependent variable becomes

$$y_{it} - \lambda\overline{y}_i$$

and the k-th independent variable

$$x_{itk} - \lambda\overline{x}_{ik}$$

with

$$\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

$$\overline{x}_{ik} = \frac{1}{T}\sum_{t=1}^{T} x_{itk}$$

Note that we almost subtract the time mean for i if λ is close to 1, i.e. if σ_ε^2 is small relative to $T\sigma_\alpha^2$.

The random effects estimator of β is ($\Omega^{-1} = \Omega^{-1/2}\Omega^{-1/2}$)

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'\Omega^{-1}y_i\right)$$

Substitution of the model gives

$$\hat{\beta} = \beta + \left(\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'\Omega^{-1}(\alpha_i\iota_T + \varepsilon_i)\right)$$

Note that for consistency we need

$$\mathrm{E}(X_i'\Omega^{-1}\varepsilon_i|X_i) = X_i'\mathrm{E}(\Omega^{-1}\varepsilon_i|X_i) = 0$$

This is equivalent to

$$\mathrm{E}(\Omega^{-1}\varepsilon_i|x_{i1},\dots x_{iT}) = 0$$

Because Ω^{-1} is not diagonal the T vector $\Omega^{-1}\varepsilon_i$ has components that are linear combinations of all $\varepsilon_{i1},\dots,\varepsilon_{iT}$. Hence the earlier assumption

$$\mathrm{E}(\varepsilon_{it}|x_{it}) = 0$$

does not imply that

$$\mathrm{E}(\Omega^{-1}\varepsilon_i|x_{i1},\dots x_{iT}) = 0$$

Instead we need that

$$\mathrm{E}(\varepsilon_{it}|x_{i1},\dots x_{iT}) = 0$$

If

$$\mathrm{E}(\varepsilon_{it}|x_{it}) = 0$$

we say that x is weakly exogenous, and if

$$\mathrm{E}(\varepsilon_{it}|x_{i1},\dots x_{iT}) = 0$$

we call x strictly exogenous. Note that in all these definitions mean independence can be replaced by uncorrelatedness.

Strict exogeneity excludes that x_{it} depends on $\varepsilon_{i,t-1},\dots$. If y_{it} is farm output and x_{it} is amount of fertilizer used, then a large negative ε_{it}, e.g. due to bad weather, may restrict the farmer when buying fertilizer in period $t+1$. In the twins example, the education choice of a one twin brother/sister could depend on omitted variables that are specific to the other twin.

If the strict exogeneity assumption holds the random effects (RE) estimator is consistent. The estimator is also asymptotically normal with variance matrix

$$\mathrm{Var}(\hat{\beta}) = \left(\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1}$$

and if $\mathrm{E}(\eta_i\eta_i'|X_i) \neq \sigma_\alpha^2\iota_T\iota_T' + \sigma_\varepsilon^2 I_T$, we can use the robust variance matrix

$$\mathrm{Var}(\hat{\beta}) = \left(\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'\Omega^{-1}\hat{\eta}_i\hat{\eta}_i'\Omega^{-1}X_i\right)\left(\sum_{i=1}^{n} X_i'\Omega^{-1}X_i\right)^{-1}$$

In the estimator and the expressions for the asymptotic variance we need to estimate Ω, i.e. we need to estimate σ_α^2 and σ_ε^2. Because

$$\mathrm{E}(\eta_{it}\eta_{is}) = \sigma_\alpha^2$$

we have the consistent estimators

$$\hat{\sigma}_\alpha^2 = \frac{2}{nT(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{t} \sum_{s=t+1}^{T} \hat{\eta}_{it} \hat{\eta}_{is}$$

and

$$\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_\eta^2 - \hat{\sigma}_\alpha^2$$

To use this we first estimate β by pooled OLS. We use the OLS residuals to estimate Ω.

Instead of estimating Ω using the estimates of the variance of α and ε we could use

$$\hat{\Omega} = \frac{1}{N} \sum_{i=1}^{n} \hat{\eta}_i \hat{\eta}_i'$$

This does not impose the RE structure.

Results for twins data

Data from Orley Ashenfelter and Alan Krueger, "Estimates of the Economic Return to Schooling from a New Sample of Twins", AER, 1994.

- Data collected at a twins convention.
- Observations on 149 identical twins.
- Idea is that by comparing twins we can control for genetic differences in ability and other differences.

Sample statistics

Variable	Mean	Stand. dev.
Male	.456	.500
Age	36.56	10.38
Educ 1	14.06	2.20
Educ 2	14.17	2.13
Log wage 1	2.41	.606
Log wage 2	2.34	.632

Pooled OLS

	OLS est.	Std. error	Robust std. error
Constant	.402	.245	.283
Age	.0185	.00310	.00387
Male	.187	.0646	.0784
Educ	.0862	.0149	.0181

Cross correlation EDUC and OLS residual: -.00485

Random effects

$\sigma_\eta^2 = .304$, $\sigma_\alpha^2 = .145$, $\sigma_\varepsilon^2 = .159$.

Note that about 50% of the error variance is due to omitted variables common to both twins.

	RE est.	Std. error	Robust std. error
Constant	.390	.265	.376
Age	.0185	.00376	.00457
Male	.187	.0784	.0878
Educ	.0870	.0155	.0213

Discussion

- There is no evidence of failure of strict exogeneity.
- The parameter estimates are similar between pooled OLS and RE.
- Important omitted common unobservables.
- RE does not deliver more accurate estimates. The standard errors are slightly larger than for pooled OLS.