USC, Fall 2016, Economics 513

Lecture 10: Panel data: Fixed effects

Fixed Effects

We consider the same model with an error component random error term that we considered in lecture 9

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

In this lecture we allow for the possibility that α_i and $x_{i1}, \ldots, x_{iT}$ are correlated. In that case the pooled OLS estimator and also the RE estimator are biased, because there is an omitted variable α_i that is correlated with the independent variables.

We consider estimators that estimate the partial effect β consistently even if α_i and $x_{i1}, \ldots, x_{iT}$ are correlated. One of the estimators will treat $\alpha_i, i = 1, \ldots, n$ as parameters, but it should be stressed that these are random variables that are part of the random error. The key is that we make no assumptions on the distribution of the α_i or the dependence of its distribution on $x_{i1}, \ldots, x_{iT}$.

The First-Differenced (FD) estimator

If we can eliminate α_i from the linear relation then we can avoid assumptions on its distribution and/or relation to $x_{i1}, \ldots, x_{iT}$.

An obvious transformation that eliminates α_i is to first-difference the model. The first-difference operator Δ is defined as

$$\Delta y_{it} = y_{it} - y_{i,t-1}$$

If we apply this operator to the left and right hand side of the linear relation we obtain

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta\varepsilon_{it}$$

If we define

$$\eta_{it} = \Delta\varepsilon_{it}$$

we obtain a linear relation in the differenced variables.

We estimate β by pooled OLS. This is the First-Differenced (FD) estimator.

The OLS estimator is consistent if

$$\mathrm{E}[\Delta\varepsilon_{it}|\Delta x_{it}] = 0$$

Note that is is true if

$$\mathrm{E}[\varepsilon_{it}|x_{it}, x_{i,t-1}] = 0$$

and

$$\mathrm{E}[\varepsilon_{i,t-1}|x_{it}, x_{i,t-1}] = 0$$

and the latter is the same as

$$\mathrm{E}[\varepsilon_{i,t}|x_{it}, x_{i,t+1}] = 0$$

Therefore a sufficient condition (because it implies the two conditions above) for this is that

$$\mathrm{E}(\varepsilon_{it}|x_{i,t+1}, x_{it}, x_{i,t-1}) = 0$$

This is weaker than the assumption of strict exogeneity

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

but still we cannot have that ε_{it} affects $x_{i,t+1}$.

By first differencing we lose one observation for each i: There are only $T-1$ differences.

Besides α_i we lose all variables that do not vary with t. In the twins model AGE and MALE are differenced out. We can only estimate the regression coefficients of the variables that vary with t, e.g. in the twins model only EDUC remains. In particular, the intercept of the relation is eliminated.

Should we have an intercept in the differenced relation? If the panel data are longitudinal data then an intercept in the differenced relation corresponds to a linear time trend in the relation for levels. Consider a relation with an an intercept a linear time trend and a single time-varying independent variable x_{it}

$$y_{it} = \beta_0 + \beta_1 t + \beta_2 x_{it} + \alpha_i + \varepsilon_{it}$$

First-differencing gives

$$\Delta y_{it} = \beta_1 + \beta_2 \Delta x_{it} + \eta_{it}$$

For the twins model we have an intercept in the differenced model if the intercept for twin 1 is different from that of twin 2. Here we expect the intercept in the differenced relation to be essentially 0.

OLS in the differenced relation gives the best possible (BLU) estimator of β if the η_{it} are homoskedastic and not correlated over t, e.g.

$$\mathrm{E}(\eta_{it}\eta_{i,t-1}|X_i) = 0$$

This is inconsistent with the same assumption for the ε_{it}. For this reason we may want to use the robust variance matrix for the first-differenced estimator. Define the $(T-1) \times K$ matrix

$$\Delta X_i = \begin{pmatrix} \Delta x'_{i2} \\ \vdots \\ \Delta x'_{iT} \end{pmatrix}$$

The robust variance matrix is

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

This variance is correct without any assumption on $\mathrm{E}(\eta_{it}\eta_{is}|\Delta X_i)$ for all t, s. Need not be 0 for $t \neq s$ and need not be homoskedastic.

The Fixed Effects (FE) estimator

Consider again

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

and we define now

$$\eta_{it} \equiv \alpha_i + \varepsilon_{it}$$

Let us treat α_i as a parameter to be estimated, i.e. it is the coefficient on a dummy variable that is 1 for the observations for i and 0 otherwise. If we estimate the coefficients β in a relation with dummy variables for all $i = 1, \ldots, n$ (can there also be an intercept?), then by partitioned regression we regress the OLS residuals of the regression of $y_{it}, t = 1, \ldots, T, i = 1, \ldots, n$ on these dummies on the residuals of the regression of $x_{it}, t = 1, \ldots, T, i = 1, \ldots, n$ on these dummy variables.

For instance, the residuals of the regression of $y_{it}, t = 1, \ldots, T, i = 1, \ldots, n$ on these dummies are

$$\begin{pmatrix} y_{11} - \overline{y}_1 \\ \vdots \\ y_{1T} - \overline{y}_1 \\ y_{21} - \overline{y}_2 \\ \vdots \\ y_{nT} - \overline{y}_n \end{pmatrix}$$

So we take all variables in deviation of their individual specific time mean with e.g.

$$\overline{y}_1 = \frac{1}{T}\sum_{i=1}^{T} y_{1t}$$

The matrix M_2 of lecture 2 that transforms the observations is the $nT \times nT$ block diagonal matrix

$$M_2 = \begin{pmatrix} E_T & \emptyset & \emptyset \\ \emptyset & \ddots & \emptyset \\ \emptyset & \emptyset & E_T \end{pmatrix}$$

with

$$E_T = I_T - \frac{1}{T}\iota_T \iota_T'$$

The Fixed Effects (FE) estimator is

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

This is the pooled OLS estimator using the transformed data

$$y_{it}^* \equiv y_{it} - \overline{y}_i \qquad x_{it}^* \equiv x_{it} - \overline{x}_i$$

To derive its properties we use

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

Now

$$E_T \eta_i = E_T(\alpha_i \iota_T + \varepsilon_i) = E_T \varepsilon_i$$

because $E_T \iota_T = 0$. Therefore we have the important result that the FE estimator does not depend on the individual effects $\alpha_1, \ldots, \alpha_n$ but only on ε_i

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

For consistency we need that

$$\mathrm{E}(\varepsilon_{it} | x_{it} - \overline{x}_i) = 0$$

and a sufficient condition for this is that the independent variables are strictly exogenous (use the Law of Iterated Expectations to show this)

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

From

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

we find that $\hat{\beta}$ is approximately normal with mean β and variance matrix

$$\left(\sum_{i=1}^{n} X_i' E_T X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i' E_T \mathrm{E}(\varepsilon_i \varepsilon_i' | X_i) E_T X_i\right) \left(\sum_{i=1}^{n} X_i' E_T X_i\right)^{-1}$$

The variance matrix simplifies if we assume

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

In that case the variance matrix is estimated by

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \hat{\sigma}_\varepsilon^2 \left(\sum_{i=1}^{n} X_i' E_T X_i\right)^{-1}$$

with

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T} e_{it}^2$$

and

$$e_{it} = y_{it} - x_{it}'\hat{\beta} - \hat{\alpha}_i$$

with

$$\hat{\alpha}_i = \overline{y}_i - \overline{x}_i'\hat{\beta}$$

Note that we divide by $n(T-1)$ because we estimated n coefficients on individual dummies.
Without the simplifying assumption the robust variance matrix is estimated by

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

Note that we use the OLS residuals defined above and not those of the transformed relation.

Results for twins data

For the twins data $T = 2$ and in that case the FD and FE estimators give identical estimates, because

$$y_{i1} - \overline{y}_i = y_{i1} - \frac{y_{i1} + y_{i2}}{2} = \frac{1}{2}(y_{i1} - y_{i2})$$

and

$$y_{i2} - \overline{y}_i = y_{i2} - \frac{y_{i1} + y_{i2}}{2} = \frac{1}{2}(y_{i2} - y_{i1})$$

and the same holds for the independent variables. Therefore we essentially have a relation between the first differences of the dependent and independent variables.

FE=FD estimates

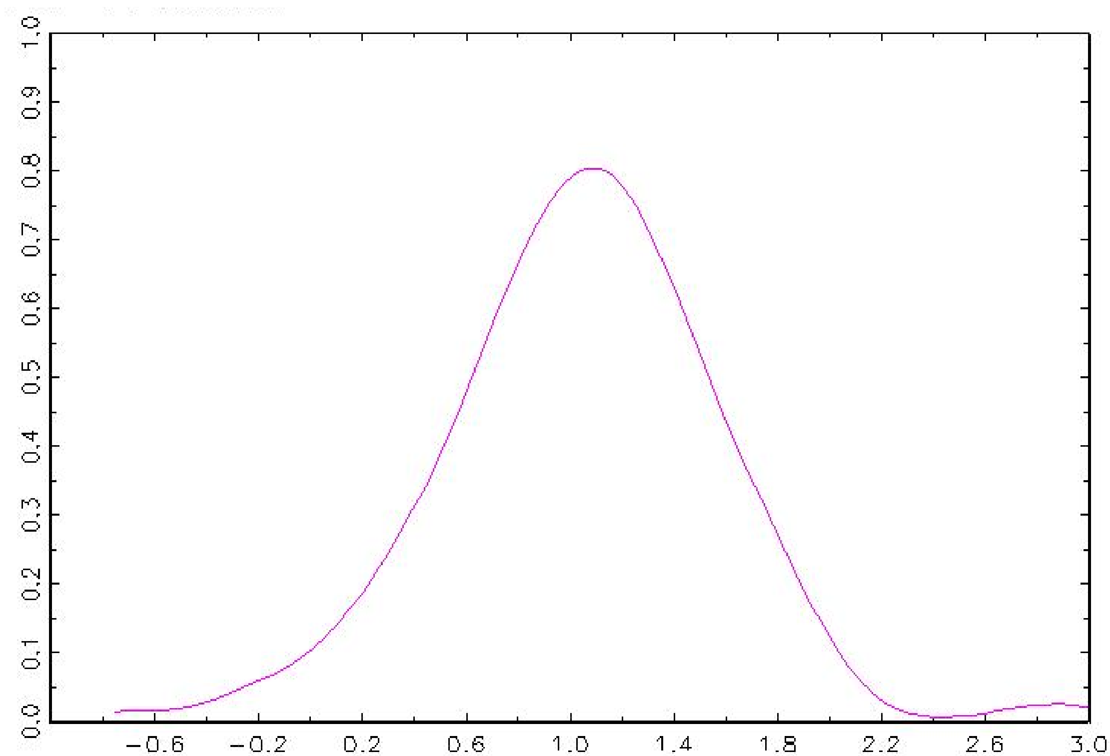
	FE est	std. err.	robust std. err.
Constant	.0786	.0455	.0448
EDUC	.0916	.0237	.0285

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

Analysis of individual effects

The individual effects have mean 1.047 and standard deviation .521. An estimate of the density is in the figure.



We estimate the linear relation between the estimates of the individual effects and AGE, the square of AGE, MALE, and the average years of education for the twins.

	OLS est.	Std. error
Const	.374	.328
AGE	.0185	.0038
MALE	.187	.0793
Av. EDUC	-.00619	.0205

We conclude that the individual effect that contains the omitted variables common to the twins, is only weakly correlated with the average level of education and this explains why the pooled OLS and RE estimators almost give the same result.