Advanced panel data methods

Lecture 2/3 EE426 – 2/2013

Chayanee Chawanote

Fixed effects estimation

Overview:

- When there is an observed fixed effect, an alternative to first differences is fixed effects estimation
- The average of a_i across t will be a_i, so if you subtract the mean, a_i will be differenced out just as when doing first differences
- This method is also identical to including a separate intercept for every individual (dummy var.)
- First Differences and Fixed Effects will be exactly the same when T = 2

Fixed effects estimation

- Consider a model with k explanatory variables:

$$y_{it} = \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + a_i + u_{it}, \quad t = 1, 2, \ldots, T \qquad (1)$$

- For each *i*, average the equation over time to get

$$\bar{y}_i = \beta_1 \bar{x}_{1i} + \ldots + \beta_k \bar{x}_{ki} + a_i + \bar{u}_i \qquad (2)$$

- Subtract (2) from (1), for each t, we have

$$y_{it} - \bar{y}_i = \beta_1 (x_{1it} - \bar{x}_{1i}) + \ldots + \beta_k (x_{kit} - \bar{x}_{ki}) + u_{it} - \bar{u}_i, \quad t = 1, 2, \ldots, T$$

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{1it} + \ldots + \beta_k \ddot{x}_{kit} + \ddot{u}_{it}, \quad t = 1, 2, \ldots, T \qquad (3)$$

- $\ddot{y}_{it} = y_{it} - \bar{y}_i$ is time-demeaned data on y
- Fixed effects estimator or within estimator: pooled OLS estimator based on the time-demeaned variables

Fixed effects estimation

- Between estimator: apply OLS estimator on the cross-sectional data equation (2) with an intercept
  - It is biased when a_i is correlated with $\overline{x}_i$
  - If we think a_i is uncorrelated with x_{it}, it is better to use the random effects estimator
  - It ignores information on how the variables change over time
- Under strict exogeneity assumption (u_{it} is uncorrelated with each x_{jit} across all time periods), FE estimator is unbiased.
  - this rules out feedback from past u_{is} shock to current x_{it}, or we cannot have lagged dependent variables on RHS
- We also need u_{it} to be homoskedastic and serially uncorrelated (across t).

Fixed effects estimation

- Degree of freedom: for each cross-sectional observation *i*, we lose one df because of the time-demeaning.
  - Use df = NT – N – k = N(T-1) – k when correcting s.e. and test
- R^2 (within transformation) from (3) is interpreted as the amount of time variation in the y_{it} that is explained by the time variation in the explanatory variables.
- We cannot include time-constant variables
  - but they can be interacted with time-varying variables, e.g., time dummy variables >> how each year differs from the base period
- We cannot estimate the effect of any variable whose change across time is constant, e.g., years of experience

Fixed effects vs. dummy variable regression

- FE is equivalent to adding dummy variables for each cross-sectional observations >> get intercept for each $i = a_i$
- But, we have N + k parameters to estimate with only N observations >> not practical
- View the a_i as omitted variables that we control for
- If the intercept is reported in FE, it is interpreted as the average across *i* of the estimated a_i (the average of the individual-specific intercepts)
  - How can we know whether unobserved fixed effect of *i* we are interested (e.g. city *i*) are above or below average?

Assumptions for pooled OLS using Fixed Effects

- FE.1: For each *i*, the model is

$$y_{it} = \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + a_i + u_{it}, \quad t = 1, \ldots, T$$

- FE.2: A random sample from the cross section
- FE.3: Each explanatory variable changes over time (for at least some i), and no perfect linear relationships exist among the explanatory variables
- FE.4: For each t, the expected value of the idiosyncratic error given the explanatory variables in all time periods and the unobserved effect is zero: $E(u_{it} | \boldsymbol{X}_i, a_i) = 0$, $\boldsymbol{X}_i$ contains x_{jit}, $t = 1, \ldots, T$; $j = 1, \ldots, k$
  - x_{jit} are strictly exogenous conditional on the unobserved effect.
  - FE estimator is also consistent with a fixed T and as N → ∞

Assumptions for pooled OLS using First Differences

- FE.5: The variance of the errors, conditional on all explanatory variables, is constant (homoskedasticity)

$$\text{Var}(u_{it}|\boldsymbol{X}_i, a_i) = \text{Var}(u_{it}) = \sigma^2_u, \text{ for all t = 2,...,T}$$

- FE.6: For all t ≠ s, the idiosyncratic errors are uncorrelated (serially uncorrelated)

$$\text{Cov}(u_{it}, u_{is}|\boldsymbol{X}_i, a_i) = 0, \text{ t} \neq \text{s}$$

  - FE.1-6: FE estimator β_j is the best linear unbiased estimator (conditional on the explanatory variables).

- FE.7: Conditional on $\boldsymbol{X}_i$ and a_i, $u_{it} \sim \text{N}iid(0, \sigma^2_u)$
  - FE estimator is normally distributed, and t and F stats have exact t and F distributions

Fixed effects (FE) vs. First Differences (FD)

- When T = 2, FE and FD are equivalent.
- Both FD and FE are unbiased (consistent) under FD.1-4 and FE.1-4, respectively. Look at the relative efficiency of the estimators.
- If u_{it} is a random walk ($u_{it} = u_{i,t-1} + \varepsilon_{it}$), then Δu_{it} is serially uncorrelated. FD will be more efficient than FE.
- Under 'classical assumptions': $u_{it} \sim iid(0, \sigma^2_u)$, FE will be more efficient.
- If $\text{cov}(u_{it}, \boldsymbol{X}_i) = 0$ but the strict exogeneity is violated (there is a lagged dependent variable), then FE likely has less bias than FD.

Random effects models

- Let the unobserved effects model is

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + a_i + u_{it} \qquad (4)$$

- where an intercept is included so that we can make the assumption that E(a_i) = 0
- Here, we think a_i is uncorrelated with each x_j in all time period: $\mathrm{Cov}(x_{jit}, a_i) = 0,\ t = 1,\ldots,T;\ j = 1,\ldots,k$
- Why don't just use a single cross section when β_j can be consistently estimated?
- Suppose we define the composite error as $v_{it} = a_i + u_i$.

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + v_{it} \qquad (5)$$

  - Because a_i is in the composite error in each time period, v_{it} are serially correlated across time.

Random effects models

- Under the random effects assumption, t ≠ s

$$corr(v_{it}, v_{is}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2}$$

- The pooled OLS standard errors ignore this correlation. They will be incorrect, so as the test statistics.
- We can use GLS to solve the serial correlation problem.
  - need large N and relatively small T
- Random Effects (RE) estimator is a GLS estimator that take the above correlation into account.

Random effects models

- Define $\lambda = 1 - \left(\frac{\sigma_a^2}{T\sigma_a^2 + \sigma_u^2} \right)^{1/2}$

- Multiply λ by the individual average of the original equation and subtract from the original equation:

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda) + \beta_1(x_{1it} - \lambda\bar{x}_{1i}) + \ldots + \beta_k(x_{kit} - \lambda\bar{x}_{ki}) + (v_{it} - \lambda\bar{v}_i) \quad (6)$$

- Then, using OLS on this transformed equation gives the random effects GLS estimator
- Now, $v_{it} - \lambda\bar{v}_i$ is serially uncorrelated.
- Since λ is unknown, we need to estimate it from estimated σ^2_a *and* σ^2_u:
  - Use pooled OLS to obtain estimates of the composite residual $\hat{v}_{it}$. Calculate σ^2_a as cov($\hat{v}_{it}, \hat{v}_{i,t-1}$), and $\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_a^2$

Random effects models

- The transformation in eq.(6) allows for explanatory variables that are constant over time.
- Relationship between RE, pooled OLS and FE from eq.(6):
  - Pooled OLS: $\lambda = 0$ (or $\hat{\lambda}$ close to zero), when the unobserved effect, a_i, is relatively unimportant because it has small variance relative to σ^2_u
  - FE: $\lambda = 1$ (or $\hat{\lambda}$ close to unity), when σ^2_a is large relative to σ^2_u
- Rewrite $v_{it} - \lambda\bar{v}_i$ as $(1-\lambda)a_i + u_{it} - \lambda\bar{u}_i$. We can see that the errors in the transformed equation in RE weight the unobserved effect by $(1-\lambda)$.

Assumptions for Random Effects

- FE.1 & FE.2
- RE.3: No perfect linear relationships exist among the explanatory variables
- RE.4: In addition to FE.4, the expected value of a_i given all explanatory variables is constant, $E(a_i|\boldsymbol{X}_i) = \beta_0$
  - to rule out correlation between the unobserved effect and the explanatory variables.
- RE.5: In addition to FE.5, the variance of a_i given all explanatory variables is constant, $Var(a_i|\boldsymbol{X}_i) = \sigma^2_a$
- FE.6: Serially uncorrelated
- Under these assumptions, RE estimators are asymptotically efficient: in large samples, RE will have smaller s.e. than pooled OLS; more efficient than FE for coefficients on time-varying explanatory variables

Random effects or Fixed effects?

- The FE estimator is always consistent, but inefficient under the null hypothesis that Cov(x_{it} , a_i) = 0. RE is both consistent and relatively efficient under that null hypothesis, but inconsistent under the alternative.
- Hausman test: the preferred model is random effects vs. the alternative the fixed effects

$$H = (\hat{\beta}^{FE} - \hat{\beta}^{RE})'[\mathrm{var}(\hat{\beta}^{FE}) - \mathrm{var}(\hat{\beta}^{RE})]^{-1}(\hat{\beta}^{FE} - \hat{\beta}^{RE}) \sim \chi^2_M$$

```
xtreg depvar indepvars1, fe
estimates store fe
xtreg depvar indepvars2, re
estimates store re
hausman fe re
```

Comparing all models

FD	FE	RE	Pooled OLS
$Cov(x_{it}, a_i) \neq 0$	$Cov(x_{it}, a_i) \neq 0$	$Cov(x_{it}, a_i) = 0$	$Cov(x_{it}, a_i) = 0$
Time-invariant vars: drop	drop	allow	allow
Efficient estimators FD vs. FE: when u_{it} is a random walk	$u_{it} \sim iid(0, \sigma^2_u)$		
	FE vs. RE: Hausman test reject H_0	cannot reject H_0	
		RE vs. Pooled OLS: Breusch-Pagan test reject H_0	cannot reject H_0

Testing for the presence of an unobserved effect: The Breusch-Pagan test

- If the regressors are strictly exogenous and u_{it} is non-autocorrelated and homoskedastic, then both pooled OLS and RE will both be efficient if no unobserved effects (σ^2_a=0). If $\sigma^2_a > 0$, then RE is efficient.
- Lagrange multiplier test >> H_0: $\sigma^2_a = 0$

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_i\left(\sum_t \hat{v}_{it}\right)^2}{\sum_i\sum_t \hat{v}_{it}^2} - 1\right]^2 \sim \chi_1^2$$

where $\hat{v}_{it}$ is the estimated pooled OLS residual

- `xttest0`: if failed to reject the null = no significance across units (no panel effect)

Applying Panel Data Methods to Other Data Structures

- FD, FE, and RE can be applied to data structures that do not involve time, e.g., demography, to control for unobserved family and background characteristics (family fixed effects)
- Cluster sample: cross-sectional data set, but each observation belongs to a well-defined cluster, e.g. each family is a cluster.
- Outcomes within a cluster are likely to be correlated, allowing for an unobserved cluster effect – use FE
- If the cluster effect is uncorrelated with all Xs – use RE
- If using pooled OLS and there is correlation within clusters, correct for s.e. – STATA option for vce cluster()