Advanced Microeconometrics 6SSPP393

Topic 3: Advanced Methods for Panel Data

Yonatan Berman and Elisa Cavatorta

Semester 2, 2023/24

Consistency and efficiency

- Consistency and efficiency are asymptotic properties
- Consistency: if an estimator is consistent, then the distribution of $\hat{\theta}$ becomes more and more concentrated about the true parameter θ as the sample size grows
  - Unlike unbiasedness- which is a feature of an estimator for a given sample size
- A good example of a consistent estimator is the average of a random sample drawn from a population with mean μ and variance σ^2
  - The variance of the sample is σ^2/n – as n increases, the variance tends to zero, so the estimator is consistent (and also unbiased)

Consistency and efficiency

- Asymptotic efficiency means that an estimator has a smaller variance as the sample size grows relative to another estimator (in the same class) – this is a relative property
- So far we mainly considered issues that could create a bias in the estimates
- For inference it is an advantage for an estimator to also be more efficient, *i.e.*, estimate a coefficient with smaller error than others
- We will address various types of estimators

Variation in the data

- We saw have seen that differencing is one method of eliminating the time-invariant unobserved heterogeneity: a_i (which itself is sometimes called a fixed effect)
- FD and the difference-in-differences estimator were based on the first-differenced transformation of the data
- Think about the variation one has in general panel data ($T > 2$): we have a time variation over t and a cross-sectional variation over i
- For each i: a_i appears in the same 'amount' every period and in the within i time average

Variation in the data

Unit i	period t	x_{it}	a_i
1	1	x_{11}	a_1
1	2	x_{12}	a_1
1	3	x_{13}	a_1
...	...	...	...
2	1	x_{21}	a_2
2	2	x_{22}	a_2
2	3	x_{23}	a_2
...	...	...	...

The "within" transformation in the data

- The "fixed effects" or "within" transformation removes a_i by subtracting the within i time averages from the i observations in each period

- In a simple model with only one x_{it}:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

- We average over t to get

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_i \ ,$$

where $\bar{y}_i = \left(\sum_{t=1}^{T} y_{it} \right) / T$ is a time average for each unit i

The "within" transformation in the data

- We can now subtract $y_{it} - \bar{y}_i$ to obtain:

$$y_{it} - \bar{y}_i = \beta_1 (x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i)$$

- As with the FD equation, this equation is free of a_i
- This is the time-demeaned equation
- We can now use pooled OLS on the deviations from the time averages to estimate β_1
- This is the "within" estimator, or fixed-effects estimator (FE)

An alternative method

- An alternative method would be a way to estimate the fixed effects
- We keep the original data (y_{it} and x_{it}) and add N dummy variables for each unit i:
  - The dummy $A^j_{it} = 1$ if $i = j$ and 0 otherwise
- We can now regress: $y_{it} = \beta_0 + \beta_1 x_{it} + \sum_{j=1}^{N} \theta_j A^j_{it} + u_{it}$
- This is called the "dummy variable" (DV) regression. It gives the same standard errors and statistics as the FE regression
- The advantage is that we also estimate this way the fixed effects, which could be of interest

Have we solved the omitted variable problem?

- In general it is better to think about a_i as unobserved factors that need to be taken into account so that no omitted variable bias is present
- Like FD, FE estimation allows arbitrary correlation between x_{it} and a_i, but it requires a strict exogeneity assumption with respect to the error:

$$Cov\left(x_{is}, u_{it}\right) = 0 \ \forall s, t$$

- This means that the error u_{it} should be uncorrelated with each explanatory variable across all time periods
- We still need to check for serial correlation (heteroskedasticity) in the errors

FE and DV estimation in Stata

- The command for FE estimation in Stata is `xtreg`
- It computes proper standard errors and test statistics, but the default is that the errors u_{it} are homoskedastic and serially uncorrelated, so this needs to be tested for and clustered if needed
- The command for DV estimation in Stata is `areg`
- Note that `xtreg` reports an intercept – there was no intercept in the FE equation! The output intercept is the average if the fixed effects
- The F statistic reported at the bottom of the `xtreg` (and `areg`) output (but only when "cluster" is not used) is a test of whether the N intercepts are all the same – its outcome is usually known a priori: a strong rejection is expected in most cases if unobserved heterogeneity is relevant

FE in Stata

```
. xtreg lscrap grant grant_1 d88 d89, fe

Fixed-effects (within) regression               Number of obs      =       162
Group variable: fcode                           Number of groups   =        54

R-sq:  within  = 0.2010                         Obs per group: min =         3
between = 0.0079                                               avg =       3.0
overall = 0.0068                                               max =         3

F(4,104)        =     6.54
corr(u_i, Xb)   = -0.0714                       Prob > F           =    0.0001

------------------------------------------------------------------------------
lscrap  |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
grant   |  -.2523149    .150629    -1.68   0.097    -.5510178    .0463881
grant_1 |  -.4215895      .2102    -2.01   0.047    -.8384239   -.0047551
d88     |  -.0802157   .1094751    -0.73   0.465     -.297309    .1368776
d89     |  -.2472028   .1332183    -1.86   0.066    -.5113797    .0169741
_cons   |   .5974341   .0677344     8.82   0.000     .4631142    .7317539
-------------+----------------------------------------------------------------
sigma_u |   1.438982
sigma_e |  .49774421
rho     |  .89313867   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(53, 104) =    24.66            Prob > F = 0.0000
```

- Note: `reghdfe` is based on `areg` (allowing for multiple fixed effects)

```
. areg lscrap grant grant_1 d88 d89, absorb(fcode)

Linear regression, absorbing indicators          Number of obs    =        162
F(    4,      104) =         6.54
Prob > F           =       0.0001
R-squared          =       0.9276
Adj R-squared      =       0.8879
Root MSE           =       0.4977

------------------------------------------------------------------------------
lscrap   |       Coef.    Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
grant    |  -.2523149     .150629     -1.68    0.097    -.5510178    .0463881
grant_1  |  -.4215895       .2102     -2.01    0.047    -.8384239   -.0047551
d88      |  -.0802157    .1094751     -0.73    0.465     -.297309    .1368776
d89      |  -.2472028    .1332183     -1.86    0.066    -.5113797    .0169741
_cons    |   .5974341    .0677344      8.82    0.000     .4631142    .7317539
-------------+----------------------------------------------------------------
fcode    |         F(53, 104) =      24.661   0.000           (54 categories)
```

POLS, FD, or FE

- How should we choose the right estimator?
- With $T = 2$, FD and FE are basically identical
- If we include an intercept in the FD equation (=which is the intercept for the second time period in the model in level), the FE must include a dummy for the second time period to be identical to the FD estimates
- For $T > 2$, FD and FE are different
- Typically find larger differences for larger T. Which one to use?

POLS, FD, or FE

- FD and FE are both unbiased if the assumptions hold

- FD and FE are both consistent for T fixed and $N \to \infty$ if the assumptions hold

- Relative efficiency depends on the serial correlation in the errors u_{it}:
  - **When N is large and T is small (*e.g.*, surveys)**:
    - When u_{it} are serially uncorrelated, FE is more efficient (why?)
    - When u_{it} are highly serially correlated, FD is more efficient (because Δu_{it} are serially uncorrelated)
    - In between – more difficult to compare and one could compare between FD and FE and check the sensitivity of the results

  - **When T is large and N is small (*e.g.*, macro indicators):**
    - If high serial correlation in x_{it} or u_{it}, FD reduces it
    - When strict exogeneity fails, FE has an advantage over FD: the bias in FE decreases at a rate $1/T$

POLS, FD, or FE

- Computing pooled OLS, FD, and FE estimators can be informative
- If FD and FE are very different, it is a sign that strict exogeneity fails
- If POLS is different from FE, it indicates explanatory variables correlated with a_i
- Goodness-of-fit with FE and DV regressions:
  - With FE, the "within" R-squared is probably most informative (based on time-demeaned equation)
  - By contrast, the usual R-squared of the DV regression is typically very high because the fixed effects count in the explanatory power – use the "within" R-squared

What if a_i is uncorrelated with x_{it}?

- Suppose we have the same equation as before (with different notation):

$$y_{it} = \delta_t + \mathbf{x}_{it}\beta + a_i + u_{it}$$

  where $\mathbf{x}_{it}\beta = \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots + \beta_k x_{itk}$

- The δ_t represents difference time intercepts (or time fixed effects)
- If our interest is in β and x_{itk} and a_i are uncorrelated for k and t, then there is no omitted variable bias
- But we know that if we leave a_i in the error term, the composite errors are serially correlated over time

What if a_i is uncorrelated with x_{it}?

- Random Effects (RE) estimator leaves the a_i in the error term and deals with the serial correlation by a transformation of the data
- The core idea of the RE estimation (and also its Achilles' heel) relies on ruling out the correlation between a_i and the explanatory variables:

$$Cov\left(x_{itj}, a_i\right) = 0, \quad j = 1, \ldots, k$$

and also strict exogeneity with respect to the errors:

$$Cov\left(x_{isj}, u_{it}\right) = 0, \quad s, t = 1, \ldots, T$$

What if a_i is uncorrelated with x_{it}?

- These imply an assumption about the composite error term: $v_{it} = a_i + u_{it}$ is uncorrelated with the explanatory variables in all time periods (like assumed in POLS)
- Leaving a_i in the error terms makes v_{it} and v_{is} serially correlated across time so this problem remains
- Another way to think about the a_i assumed by the RE estimation is that they are realisations of a random variable in a random sample from a larger population of all i units and its expected value is a constant: $E\left[a_i|\mathbf{x}_{it}\right] = \beta_0$

A closer look at the RE composite error

- Let's unpack the error $v_{it} = a_i + u_{it}$
- For all random draws i from the population, the RE estimation assumes:

$$
\begin{aligned}
Cov\left(a_i, u_{it}\right) &= 0 \\
Var\left(a_i\right) &= \sigma_a^2 \text{ fixed variance across } i \\
Var\left(u_{it}\right) &= \sigma_u^2 \text{ fixed variance over time} \\
Cov\left(a_i, u_{it}\right) &= 0, t \neq s \text{ no serial correlation}
\end{aligned}
$$

- In practice, there could be serial correlation, and again – we could cluster errors to make sure the errors are robust

The serial correlation in the composite error

- The RE estimation use properties of the composite error, $v_{it} = a_i + u_{it}$:

$$\begin{aligned} Var\left(v_{it}\right) &= \sigma_a^2 + \sigma_u^2 \\ Cov\left(v_{it}, v_{is}\right) &= Var\left(a_i\right) = \sigma_a^2 \end{aligned}$$

  where the first follows from $Cov\left(a_i, u_{it}\right) = 0$ and the last follows from $Cov\left(u_{it}, u_{is}\right) = 0,\ t \neq s$

- The serial correlation in v_{it} over time is:

$$Corr\left(v_{it}, v_{is}\right) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \equiv \rho$$

- ρ is the share of the total variance $\sigma_a^2 + \sigma_u^2$ due to the variance in a_i, the unobserved effect, σ_a^2

RE estimation: transforming the data

- The RE estimator also transforms the data, like the FE estimator, but uses a different transformation
- We define a parameter θ that is between 0 and 1:

$$\theta = 1 - \left[\frac{1}{1 + T\left(\sigma_a^2/\sigma_u^2\right)}\right]^{1/2}$$

- The RE estimate can be obtained from the pooled OLS regression:

$$y_{it} - \theta\bar{y}_i = \beta_0\left(1-\theta\right) + \beta_1\left(x_{it} - \theta\bar{x}_{1i}\right) + \ldots + \left(v_{it} - \theta\bar{v}_i\right)$$

- $y_{it} - \theta\bar{y}_i$ is a partially-time-demeaned variable
- To estimate the variances σ_a^2 and σ_u^2 using pooled OLS or FE estimation (we get in fact $\hat{\theta}$)

Partially demeaned data, FE and POLS

- We defined

$$
\begin{aligned}
\theta &= 1 - \left[\frac{1}{1 + T\left(\sigma_a^2/\sigma_u^2\right)}\right]^{1/2} \\
y_{it} - \theta \bar{y}_i &= \beta_0 \left(1 - \theta\right) + \beta_1 \left(x_{it} - \theta \bar{x}_{1i}\right) + \ldots + \left(v_{it} - \theta \bar{v}_i\right)
\end{aligned}
$$

- It is easy to see that

$$
\begin{aligned}
\hat{\theta} &\approx 0 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{POLS} \\
\hat{\theta} &\approx 1 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{FE}
\end{aligned}
$$

- When is $\hat{\theta} \approx 1$?
  I. $\sigma_a^2/\sigma_u^2 \gg 1$ or
  II. T is large

RE regression in Stata

- No need to create the transformation by hand – there is a stata command:

```
xtset id year
xtreg y x1 x2 ...  xK, re
xtreg y x1 x2 ...  xK, re cluster(id)
```

Takeaway points

- We have seen four ways to deal with time-invariant unobserved heterogeneity (a_i): Pooled OLS, FD, FE and RE
- Comparing POLS, FD, RE, and FE is informative of the bias caused by leaving a_i in the error term:
  - Entirely in POLS
  - partially in RE
  - Removed in FE and FD
- POLS, RE and FE are all consistent if $Cov\left(a_i, x_{it}\right) = 0$
- RE is typically more efficient, but if $Cov\left(a_i, x_{it}\right) \neq 0$, it is not consistent
- If the a_i themselves are of interest, it is possible to use Dummy Variable regression

For next week

- Revise JW Chapter 14
- Example 14.4 in JW Chapter 14