TA Session 1: Panel Data

Microeconometrics I with Joan Llull

IDEA, Fall 2024

TA: Conghan Zheng

September 20, 2024

TA sessions

- One week before the due date of each problem set, we will review exercises that prepare you for it. We will also discuss the solutions and common mistakes from the previous problem set.
- Office hours: upon request.

Problem sets

- There will be four problem sets, each corresponding to chapters 2 through 5.
- Similar sentences in the interpretations of results will cause you a huge grade loss. Creative and insightful interpretations and discussions will earn you a bonus in the grade of that solution.
- You may get answers from past cohorts, but I suggest you don't look at the answers when you are solving the PS to avoid any anchoring effects. I'm the one who wrote the answers, I will recognize similar interpretations.
- This course recommends you use Stata or Matlab. For solutions I also accept R, Python, and Julia (these are all the languages I can give you feedback on).
- When you get stuck on anything in the problem sets, contact me via email: conghan.zheng@uab.cat

Overview

Model

- In this session, we are going to estimate the following model using data from British firms (see `TA1.dta`):

$$n_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 w_{it} + \beta_3 y_{it} + (u_i + e_{it})$$

where

n_{it}: firm i's employment at t
k_{it}: log capital
w_{it}: log wages paid
y_{it}: output
u_i and e_{it}: unobserved heterogeneity and idiosyncratic error

Manipulating Panel Data

Reshaping data

Original wide data (TA2.dta)

	firm	n1976	n1977	n1978	n1979	n1980	n1981	n1982	n1983	n1984
1	1	.	1.6176045	1.7227666	1.6124334	1.5507489	1.4092782	1.1524689	1.0770481	.
2	2	.	4.2671628	4.257639	4.2615243	4.2770966	4.2998536	4.2824687	4.2270965	.
3	3	.	2.952616	2.9673328	2.9907197	3.0076608	2.9739978	2.8972922	2.8243507	.
4	4	.	3.2642315	3.2861606	3.3061539	3.3261146	3.3020766	3.1988364	3.1162671	.
5	5	4.4621886	4.4670569	4.4659081	4.5042443	4.490881	4.4152196	4.3000028	.	.
6	6	-.29035227	-.26657314	-.27180869	-.31608158	-.31334181	-.24974425	-.24590053	.	.
7	7	.47000364	.50077527	.51879376	.51879376	.50681758	.44468578	.43178239	.	.
8	8	2.2132073	2.3846258	2.3748127	2.394161	2.3700568	2.1009584	1.5477753	.	.
9	9	.69614271	.97682119	.99768633	1.0094167	1.0141431	1.046968	1.012328	.	.
10	10	1.3410354	1.3699109	1.2607312	1.258461	1.2432897	1.1281711	1.1823405	.	.

Reshaping data

```
. reshape long n w k y, i(firm) j(year)
```

	firm	year	n	w	k	y
1	1	1976	.	.	.	.
2	1	1977	1.6176045	2.5765434	-.52865022	4.5612935
3	1	1978	1.7227666	2.5097456	-.4591824	4.5783836
4	1	1979	1.6124334	2.5525264	-.3899363	4.6012455
5	1	1980	1.5507489	2.6249511	-.48272419	4.6106561
6	1	1981	1.4092782	2.659539	-.67806152	4.6007414
7	1	1982	1.1524689	2.699218	-.86061956	4.5912244
8	1	1983	1.0770481	2.6231022	-.93649347	4.6054711
9	1	1984	.	.	.	.
10	2	1976	.	.	.	.
11	2	1977	4.2671628	2.6940121	2.8294593	4.5612935
12	2	1978	4.257639	2.6464301	2.8473599	4.5783836
13	2	1979	4.2615243	2.7049387	2.8645581	4.6012455
14	2	1980	4.2770966	2.7402592	2.871155	4.6106561
15	2	1981	4.2998536	2.7848198	2.8162049	4.6007414
16	2	1982	4.2824687	2.7807676	2.7879022	4.5912244
17	2	1983	4.2270965	2.7914779	2.8547216	4.6054711
18	2	1984	.	.	.	.

Drop missing

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
firm	1,260	70.5	40.42953	1	140
year	1,260	1980	2.583014	1976	1984
n	1,031	1.056002	1.341506	-2.263364	4.687321
w	1,031	3.142988	.2630081	2.081577	3.8118
k	1,031	-.4415775	1.514132	-4.431217	3.852441
y	1,031	4.638015	.0939612	4.464758	4.85488

- Some firm-year combinations are not observed.

Drop missing

```
. drop if n==. | w==. | k==. | y==.
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
firm	1,031	73.20369	41.23333	1	140
year	1,031	1979.651	2.21607	1976	1984
n	1,031	1.056002	1.341506	-2.263364	4.687321
w	1,031	3.142988	.2630081	2.081577	3.8118
k	1,031	-.4415775	1.514132	-4.431217	3.852441
y	1,031	4.638015	.0939612	4.464758	4.85488

- Now different variables have the same number of observations. Although the panel is still *unbalanced* (not the same number of observations for each year).

```
. tab year
```

year	Freq.	Percent	Cum.
1976	80	7.76	7.76
1977	138	13.39	21.14
1978	140	13.58	34.72
1979	140	13.58	48.30
1980	140	13.58	61.88
1981	140	13.58	75.46
1982	140	13.58	89.04
1983	78	7.57	96.61
1984	35	3.39	100.00
Total	1,031	100.00	

Exclude single observations and duplicates

- To exclude firms that only appear once in the panel (who can't provide time variation), we generate a variable that counts the number of times each firm appears.

```
.   by firm:  gen count = _N
.   drop if count == 1
```

- Stata interprets `_N` as the total number of observations in the by-group and `_n` to be the observation number within the by-group.
- Keep first appearances of duplicates:

```
.   bysort firm year:  keep if _n==1
```

Weights

- Why we use weights?
  - If under some sampling design, some observations are more likely to be sampled, then the statistics computed from the data are not representative of the population.
  - The sampling weights are often reported in survey data to obtain means in the presence of missing data.
  - **Sampling weights** (Stata's `pweight`) denote the inverse of the probability that the observation is included because of the sampling design.

Panel Structure

- First, set the panel data structure, which allows us to use panel data operators. The syntax is `xtset idvar timevar`.

```
. xtset firm year
```

- Have a look at the structure of our unbalanced panel:

```
. xtdescribe
```

```
    firm:  1, 2, ..., 140                                  n =        140
    year:  1976, 1977, ..., 1984                           T =          9
           Delta(year) = 1 unit
           Span(year)  = 9 periods
           (firm*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         7       7       7         7         8       9       9

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+-----------
       62     44.29   44.29 |  1111111..
       39     27.86   72.14 |  .1111111.
       19     13.57   85.71 |  .11111111
       14     10.00   95.71 |  111111111
        4      2.86   98.57 |  11111111.
        2      1.43  100.00 |  ..1111111
 ---------------------------+-----------
      140    100.00         |  XXXXXXXXX
```

Static Models

Static Models

- We will address the following questions using statistics from regression tables and tests:
  1. Does the panel structure matter? If not, a pooled regression (OLS) is sufficient.
  2. If the panel structure matters, should we use FE or RE?
  3. If FE is chosen, are one-way fixed effects sufficient, or should we include time fixed effects as well?

Fixed Effects Model

```
. xtreg n k w y, fe

Fixed-effects (within) regression               Number of obs     =      1,031
Group variable: firm                            Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6143                                         min =          7
     Between = 0.8483                                         avg =        7.4
     Overall = 0.8348                                         max =          9

                                                F(3,888)          =     471.39
corr(u_i, Xb) = 0.5926                          Prob > F          =     0.0000
```

n	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
k	.5489458	.0211507	25.95	0.000	.5074346	.590457
w	-.3106426	.0499301	-6.22	0.000	-.4086374	-.2126479
y	.5370106	.0534193	10.05	0.000	.4321679	.6418533
_cons	-.2159125	.3108411	-0.69	0.487	-.8259814	.3941565
sigma_u	.66133383					
sigma_e	.13015331					
rho	.96271231	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(139, 888) = 123.02                  Prob > F = 0.0000
```

- The estimate for _cons: $\bar{u}_i = \frac{1}{N}\sum_{i=1}^{N} u_i \approx -0.216$
- A constant term after within transformation? Stata actually fits $(y_{it} - \bar{y}_i + \bar{y}) = \bar{u} + (x_{it} - \bar{x}_i + \bar{x})'\beta + (e_{it} - \bar{e}_i + \bar{e})$

Fixed Effects Model

```
. xtreg n k w y, fe

Fixed-effects (within) regression               Number of obs     =      1,031
Group variable: firm                            Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6143                                         min =          7
     Between = 0.8483                                         avg =        7.4
     Overall = 0.8348                                         max =          9

                                                F(3,888)          =     471.39
corr(u_i, Xb) = 0.5926                          Prob > F          =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |   .5489458   .0211507    25.95   0.000     .5074346     .590457
           w |  -.3106426   .0499301    -6.22   0.000    -.4086374   -.2126479
           y |   .5370106   .0534193    10.05   0.000     .4321679    .6418533
       _cons |  -.2159125   .3108411    -0.69   0.487    -.8259814    .3941565
-------------+----------------------------------------------------------------
     sigma_u |  .66133383
     sigma_e |  .13015331
         rho |  .96271231   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(139, 888) = 123.02                  Prob > F = 0.0000
```

- The estimate for $\texttt{rho} = \frac{\texttt{sigma\_u}^2}{\texttt{sigma\_u}^2 + \texttt{sigma\_e}^2} = \frac{\widehat{Var}(u_i)}{\widehat{Var}(u_i) + \widehat{Var}(e_{it})} \approx 0.963$

Fixed Effects Model

```
. xtreg n k w y, fe

Fixed-effects (within) regression               Number of obs     =      1,031
Group variable: firm                            Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6143                                         min =          7
     Between = 0.8483                                         avg =        7.4
     Overall = 0.8348                                         max =          9

                                                F(3,888)          =     471.39
corr(u_i, Xb) = 0.5926                          Prob > F          =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |   .5489458   .0211507    25.95   0.000     .5074346     .590457
           w |  -.3106426   .0499301    -6.22   0.000    -.4086374   -.2126479
           y |   .5370106   .0534193    10.05   0.000     .4321679    .6418533
       _cons |  -.2159125   .3108411    -0.69   0.487    -.8259814    .3941565
-------------+----------------------------------------------------------------
     sigma_u |  .66133383
     sigma_e |  .13015331
         rho |  .96271231   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(139, 888) = 123.02                  Prob > F = 0.0000
```

- `F test that all u_i=0`: we shall reject the null from the fact `Prob > F = 0.0000`, an FE estimation is better than a pooled regression.

Random Effects Model (FGLS estimator)

- Now we have confirmed the existence of individual-specific effects. We still want to check the performance of an RE estimation.
- Is RE also better than a pooled regression? If not, then it will be kicked out of our options. Otherwise, we still need to consider the essential question for panel data: RE or FE?
- LM test on $H_0 : \sigma_u^2 = 0$

```
Breusch and Pagan Lagrangian multiplier test for random effects

        n[firm,t] = Xb + u[firm] + e[firm,t]

        Estimated results:
                         |       Var     SD = sqrt(Var)
                ---------+-----------------------------
                       n |    1.79964       1.341506
                       e |   .0169399       .1301533
                       u |   .2747343       .5241511

        Test: Var(u) = 0
                             chibar2(01) =   3044.54
                          Prob > chibar2 =    0.0000
```

`Prob > chibar2` tells us that we shall reject the null and choose RE over a pooled regression.

How to read R^2

```
Fixed-effects (within) regression
Group variable: firm

R-squared:
     Within  = 0.6143
     Between = 0.8483
     Overall = 0.8348
```

- R^2 is the squared correlation between that actual and fitted values of the dependent variable, i.e., the fraction of the variation in y explained by $\hat{y}$.
- Three R^2 measures are provided in the table. Not all of them have the properties of a linear regression R^2.

$$\begin{aligned} \text{within } R^2 : &\quad corr^2\{(y_{it} - \bar{y}_i), (x_{it} - \bar{x}_i)'\hat{\beta}\} \\ \text{between } R^2 : &\quad corr^2\{\bar{y}_i, \bar{x}_i'\hat{\beta}\} \\ \text{overall } R^2 : &\quad corr^2\{y_{it}, x_{it}'\hat{\beta}\} \end{aligned}$$

Static Models

- How could a within estimator best explain between variation? (0.8483>0.6143 on last page)
- Let's have a overview of the variance decomposition:

```
.  xtsum n k w y
```

Variable		Mean	Std. Dev.	Min	Max	Observations
n	overall	1.056002	1.341506	-2.263364	4.687321	N = 1031
	between		1.33915	-2.043388	4.62618	n = 140
	within		.1945829	.242462	2.148389	T-bar = 7.36429
k	overall	-.4415775	1.514132	-4.431217	3.852441	N = 1031
	between		1.509255	-3.707425	3.625241	n = 140
	within		.2159628	-1.201903	.7753638	T-bar = 7.36429
w	overall	3.142988	.2630081	2.081577	3.8118	N = 1031
	between		.2439766	2.163658	3.583478	n = 140
	within		.0831353	2.427872	3.653328	T-bar = 7.36429
y	overall	4.638015	.0939612	4.464758	4.85488	N = 1031
	between		.0393496	4.564672	4.738875	n = 140
	within		.0855069	4.472483	4.79896	T-bar = 7.36429

- For all variables but y, there is more variation across individuals (larger between-s.d.) than over time (within-variation), so within-estimation (FE) will lead to a efficiency loss on these variables.

Least-squares dummy-variable (LDSV) estimator

- For linear models, the within estimator (FE) is equivalent to the Least-squares dummy-variable (LDSV) estimator. The LDSV model introduces N individual specific dummies $d_{j,it}$ which capture the individual fixed effects:

$$y_{it} = x_{it}'\beta + \sum_{j=1}^{N} \alpha_i d_{j,it} + \varepsilon_{it} \qquad (1)$$

  where $d_{j,it}$ is equal to 1 if $j = i$ and zero otherwise.
- Direct estimation of (1) is computationally expensive because N more regressors are introduced into the estimation, and usually we don't need the estimate of every single dummy[1].

[1] the `absorb` option of `areg` command solves this problem

First-Differenced (FD) estimator

- Consistent estimation of β in the FE model requires eliminating α_i. One way to do so is to first-difference, leading to the first-difference estimator. First-difference replies on weaker exogeneity assumptions, compare to mean-difference (FE estimator).

$$(y_{it} - y_{i,t-1}) = (x_{it} - x_{i,t-1})'\beta + (\varepsilon_{it} - \varepsilon_{i,t-1}) + \delta \tag{2}$$

- Notice that an intercept included in (2) implies that the model have a time trend, because $\delta t - \delta(t-1) = \delta$. Be careful with it according to the empirical content of your model.
- The FD estimator uses one less year of data compared with the within estimator. Usually we believe FE is more efficient than FD (please think about why), so FD is not used widely in practice.

Time effects

- Given FE, should we also include the time fixed effects?
- Define the time dummies $\{year1, year2, ..., year7\}$:

```
.   tab year, gen (year)
```

- Include the time indicator in an FE regression, and test their joint significance agains the null[2] $H_0:\ (year2, ..., year7)' = (0, ..., 0)'$. using command

```
.   xtreg n k w y year2-year7, fe vce(robust)
.   test year2 year3 year4 year5 year6 year7
```

```
 ( 1)  year2 = 0
 ( 2)  year3 = 0
 ( 3)  year4 = 0
 ( 4)  year5 = 0
 ( 5)  year6 = 0
 ( 6)  year7 = 0

       F(  6,   139) =    4.43
            Prob > F =    0.0004
```

As `Prob > F = 0.0004`, we shall reject the null and add time fixed effects to the estimation.

[2] notice that $year1$ is left out as a base category

RE or FE?

- Estimates with cluster-robust errors:

	(1) FE_oneway	(2) FE_twoway	(3) RE
k	0.549***	0.548***	0.639***
	(0.0489)	(0.0507)	(0.0342)
w	-0.311**	-0.297*	-0.290**
	(0.115)	(0.126)	(0.109)
y	0.537***	0.265	0.440***
	(0.102)	(0.153)	(0.0954)
N	1031	1031	1031

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

- Two-way FE has smaller estimates, because some variation is absorbed by time fixed effects.
- The std. err. for RE are smaller than those for the within estimators (FE), because RE uses both between variation and within variation.

More about standard errors

RE or FE?

- In cases where your variables X_t do not vary much over time, FE (and FD) can lead to imprecise estimates. You might be forced to use RE estimation.
- Generally, to deal with the key consideration in choosing between an RE and an FE approach, we want a test on whether u_i and X_{it} are correlated.

Case	RE estimator	FE estimator	Preferred Estimator
$\mathbb{E}(u_i X_{it}) = 0$	consistent; more efficient than FE	consistent; less efficient than RE	RE
$\mathbb{E}(u_i X_{it}) \neq 0$	not consistent;	consistent	FE

RE or FE?

Hausman Test for fixed effects

- *Hausman (1978)* proposed a test based on the difference between the RE and FE estimates.
- Main caveat of Hausman test:
  - (At least) Three assumptions (1) Strict exogeneity $\mathbb{E}(e_{it}|X_i, u_i) = 0$, (2) homoskedasticity, and (3) no-serial-correlation in e_{it} are maintained under the null and the alternative.
  - So this test is valid only if inference is based on default standard errors, never on cluster-robust errors.[3]
  - What makes the situation worse: the Hausman test has no systematic power against the alternative that (1) is true but (2) and (3) are false.
  - Because FE only identifies coeffcients on time-varying regressors (if not, they go into fixed effects), we clearly cannot compare FE and RE coeffcients on time-constant variables. And more: we cannot make comparison on regressors that change only across time (where FE and RE will deliver the same estimates) neither.

[3] You can compare the default s.e. with the cluster-robust s.e., in our case (try FE and RE with and without indicating cluster-robust error option), the difference is more than double, therefore we shouldn't use Hausman test. But for teaching purpose, I force the test...

RE or FE?

- Hausman Test for fixed effects

```
. hausman FE RE, constant sigmamore
```

	Coefficients (b) FE	(B) RE	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
k	.5489458	.639224	-.0902782	.0126434
w	-.3106426	-.2900276	-.020615	.0140536
y	.5370106	.4400793	.0969312	.0139812

```
                     b = Consistent under H0 and Ha; obtained from xtreg.
      B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

    chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =  53.08
Prob > chi2 = 0.0000
```

- The overall chi-square leads to a strong rejection of the null $(plim(\hat{\theta}_{RE} - \hat{\theta}_{FE}) = 0$, the fully efficient estimator and the always consistent estimator are similar), FE is preferred.
- Solutions to the caveats of Hausman test (optional): adopt the Wald test using cluster-robust standard errors, or bootstraping[4].

[4] See Wooldridge (2010) sections 10.7.3 and 12.8.2

Panel IV

- We are already familiar with the IV for cross-sectional data.
- For FE, first we do demeaning or first difference, and then apply 2SLS or GMM.
- For RE, first we do the feasible GLS transformation, and then apply 2SLS or GMM.
- Luckily, by specifying extra options, the command `xtivreg` can do the pre-estimation transformations automatically.

Dynamic Models

Dynamic Models

- Now we consider the usual individual-specific effects panel data model, with the complication that the regressors include the dependent variable lagged once (*AR(1)*).

$$n_{it} = \alpha n_{i,t-1} + \beta_0 + \beta_1 k_{it} + \beta_2 w_{it} + \beta_3 y_{it} + (u_i + e_{it}), \;\; t = 1, \ldots, T \quad (3)$$

$$Assumption \;\; D.(1): \quad \mathbb{E}(e_{it} | \underbrace{n_{i,t-1}, \ldots, n_{i0}}_{\equiv \text{ information set } I_t}, u_i) = 0$$

- From the fact that

$$\mathbb{E}(e_{it}|I_t, u_i) = \mathbb{E}(e_{it}|n_{i,t-1}, I_{t-1}, u_i) = 0,$$

we can say the current model has the dynamics completely specified: once we control for u_i, only one lag of n_{it} is necessary.

Anderson and Hsiao (1982)

- *Anderson and Hsiao (1982)* proposed pooled IV estimation of the FD equation

$$\Delta n_{it} = \alpha \Delta n_{i,t-1} + \Delta X_{it}\beta + \Delta e_{it}, \quad t = 2, \dots, T \tag{4}$$

with instrument $n_{i,t-2}$ or $\Delta n_{i,t-2}$.

- After apply this method to our data, the results would be consistent (theoretically) but are quite disappointing. The coefficients on differenced lagged n (i.e., estimate for α) exceeds one, a value not consistent with dynamic stability.

D.n	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
nL1						
D1.	2.307626	1.973194	1.17	0.242	-1.559762	6.175015
nL2						
D1.	-.224027	.179043	-1.25	0.211	-.5749448	.1268908

Arellano and Bond (1991): Difference GMM

- To estimate the first-difference model (4), *Arellano and Bond (1991)* proposed panel GMM estimators using all lags as instruments: $n_{i,t-2}, n_{i,t-3}, \dots$. It's more efficient as it uses more instruments but valid only if there is no serial correlation in the error term e_{it}.
- Obviously, now there are more instruments than endogenous variables, and we will apply GMM on the FD model, so this is called the *Difference GMM*.
- Shortcoming 1: weak instrument problem when ...
  - For large T (usually, we call it *large* at when $T > N$), there will be too many instruments.
  - When $\alpha \to 1$, the time variation is small, the sequence: $\{\Delta n_{i,t-2}, \Delta n_{i,t-3}, \dots\}$ is highly persistent, adding more instruments are not providing much more new information. Relevance of the instruments decreases.
- Shortcoming 2: obviously, regressors with no time variation are excluded from the estimation.

Arellano and Bond (1991): Difference GMM

Difference GMM with STATA (refer to Roodman, 2009):

- First, install the command `xtabond2` with: `ssc install xtabond2`
- Syntax: `xtabond2 n L.n w k y, gmmstyle(L.n) ivstyle(w k y) noleveleq`
- Option `gmmstyle`: specifies the endogeneous variables. We can limit the number of lags we instrument it with, using `gmmstyle(varlist, lag(a b))`, so that we instrument only with a to b lags.
- Option `ivstyle` specifies the variables that serve as instruments (no need to include lagged dependent variable).
- Do no forget the `noleveleq` option, otherwise the command performs a System GMM (see next slide).
- Notation: `L.n` is n with one lag, `L(a/b).n` is n from lag a to lag b.

Arellano and Bover (1995)/Blundell and Bond (1998): System GMM

- *Arellano and Bover (1995)* suggests going back to the estimation of the level equation (3):

$$n_{it} = \alpha n_{i,t-1} + \beta_0 + \beta_1 k_{it} + \beta_2 w_{it} + \beta_3 y_{it} + (u_i + e_{it}), \ \ t = 1, \ldots, T$$

  and using $\{\Delta n_{i,t-1}, \Delta n_{i,t-2}, \ldots\}$ as the instruments for $n_{i,t-1}$. This is then called the *Level GMM*.
- It's valid under: (1) orthogonality: $\{\Delta n_{i,t-1}, \Delta n_{i,t-2}, \ldots\}$ not correlated with u_i, (2) no serial correlation in e_{it}.
- *Blundell and Bond (1998)* combine Difference GMM and Level GMM, and estimate the system of level equation and first difference equation, therefore it's called a *System GMM*.
- It requires the union of the assumptions for Difference GMM and Level GMM.

Arellano and Bover (1995)/Blundell and Bond (1998): System GMM

- Compare to Difference GMM, system GMM can estimate the parameters of time-invariant regressors (from the level part).
- Compare to Level GMM, the extra moment conditions are especially helpful for improving the precision in the GMM estimator when α is close to one.
- Acemoglu et al. (2008) uses Difference GMM on panel data and finds no causal effect of income on democracy. Spilimbergo (2009) uses System GMM on panel data and find no evidence that foreign-educated individuals foster democracy in their home countries.
- In STATA: command `xtabond2` without option `noleveleq` which removes the level equation.

Appendix

Serial Correlated Standard Errors

- Different estimators are based on different assumptions.
- In static models, we assume i.i.d. errors. All $i \times t$ errors have the same variance:

$$\varepsilon_{it} \sim (0, \sigma_\varepsilon), \;\; \forall i, \forall t$$

- Usually it's more realistic to assume that ε_{it} are independent over i, whereas serial correlation $\mathbb{E}(\varepsilon_{it} \cdot \varepsilon_{is}), \forall s \neq t$ is very likely to happen.
- The empirical content is: observations for the same individual i are more likely to be correlated across time t, while the observations for different individuals are relatively less correlated.
- What happens if the i.i.d. errors assumption is violated, and there are serial correlated errors[5]?

[5] the following discussion could also be applied to the case of spatial correlated errors

Consistency of estimators

- Consider model

$$y_{it} = X'_{it}\beta + \underbrace{u_i + e_{it}}_{\equiv \varepsilon_{it}} \tag{5}$$

where X_{it} and β are $K \times 1$ vectors, other elements are all scalars.

- The pooled OLS estimator is still consistent when there is serial correlation in the error term ε if these assumptions hold:
  1. $\{y_i, X_i\}$ is an observable ergodic stationary process, where X_i is the $T \times K$ matrix of stacked X_{it}.
  2. Linearity: y is linear on X;
  3. Correct model specification: $\mathbb{E}(\varepsilon_i | X_i) = 0$ almost surely.
  4. Nonsingularity: the $K \times K$ matrix $\mathbb{E}(X_{it}X'_{it})$ is symmetric, finite, and nonsingular.
  5. For $j \in \{0, \pm 1, \dots\}$, the $K \times K$ long-run auto-covariance matrix of $\{X_{it}\varepsilon_{it}\}$

     $$V_{it} \equiv \sum_{j=-\infty}^{\infty} \mathbb{E}(X_{it}\varepsilon'_{it}\varepsilon_{i,t-j}X'_{i,t-j}) \tag{6}$$

     is positive definite.

Consistency of estimators

- Consider the consistency of the pooled OLS estimator of (7), recall that

$$\hat{\beta}_{POLS} - \beta = \left(\sum_{i=1}^{N} X_i' X_i \right)^{-1} \sum_{i=1}^{N} X_i' \varepsilon_i \tag{7}$$

- Under assumptions 1, 2, 4, and WLLN for an ergodic stationary process, we have

$$\frac{1}{N} \sum_{i=1}^{N} X_i' X_i \xrightarrow{p} \mathbb{E}(X_i' X_i) \tag{8}$$

- By assumptions 1, 3, 5, WLLN for an ergodic stationary process, and the law of iterated expectations, we have

$$\frac{1}{N} \sum_{i=1}^{N} X_i' \varepsilon_i \xrightarrow{p} \mathbb{E}(X_i' \varepsilon_i) = 0 \tag{9}$$

- It then follows that (7)&(8)&(9) $\Rightarrow \hat{\beta}_{POLS} - \beta = O_p(1) \cdot o_p(1) \xrightarrow{p} 0$, $\hat{\beta}_{POLS}$ is consistent.

Clustered Robust Inference

- Given consistency, what can we say about the efficiency?
- Suppose assumptions 1 to 5 hold. Then as $N \to \infty$, by CLT,

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, Q^{-1} V Q^{-1}\right)$$

  where $Q = \mathbb{E}(X_i' X_i)$ and the elements of V are defined in (6).
- Notice that the variance-covariante of ε goes into V, if we don't adjust our estimator for the serial correlation in ε of unknown form, the estimator is no longer efficient.

Heteroskedasticity-robust standard errors

- The problem of conditional heteroskedasticity can be included in the discussion of correlated errors (try to think about why).
- Heteroskedasticity is the case when $Var(\varepsilon_i|X_i)$ is a diagonal matrix, the diagonal elements of which assume at least two different values. Since all the off-diagonal elements are zero, we could treat the variance-covariance matrix as T more parameters to estimate.
- But it's impossible to estimate those T unknown diagonal elements consistently using a time series of T periods (they are only identified up to scale).
- If luckily, $K < T$ (we have less variables than the number of periods), we estimate $\sum_t X_t' e_t e_t' X_t$ which is $K \times K$, rather than estimating $Var(\varepsilon_i|X_i)$ which is $T \times T$, to save computing power.
- This is $K \times K$ estimator is called the White's (1980) heteroskedasticity-consistent variance-covariance matrix estimator, and is reported by the the STATA option `vce(robust)`[6].

[6] for `xtreg ..., fe`, specifying `vce(robust)` is equivalent to specifying `vce(cluster p)`

Serial Correlated Standard Errors

- Under serial correlated errors, the off-diagonal elements of the variance-covariance matrix are not zero, the errors reported by the default `vce` option (who doesn't adjust for anything) and `vce(robust)` (who only adjust for the diagonal elements) are both invalid.
- You need to tell STATA how the errors are correlated. For example, the observations could be clustered by villages (if spatial data), cohort (if labor analysis), or in our case, time.
- You can indicate this by option syntax `vce (cluster` *`clustvar`*`)` of the regression command. And the estimate will be both heteroskedasticity- and cluster-robust.
- Read the references for more about different estimates and tests for standard errors: bootstrap, kernel regression, Breusch-Godfrey test...

Back to Estimations Comparison

References

- Amemiya, T. (1994). Introduction to statistics and econometrics. Harvard University Press. Chapter 6.
- Arellano, M. (2003). Panel data econometrics. OUP Oxford. Chapters 7-8.
- Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press. Chapters 10-11.
- Hong, Y. (2020). Foundations of Modern Econometrics: A Unified Approach. Chapters 5-6.
- Hansen, B. E. (2022). Econometrics. Chapter 17.
- Cameron, A. C., & Trivedi, P. K. (2022). Microeconometrics using stata (Second Edition). Stata press. Chapters 8-9.