Practicing Fixed Effects Methods with Artificial Data

Yona Rubinstein

July 2016

Practice with Made Up Data

- To illustrate the concept of individual fixed effects and practice fixed effects estimation procedure let's construct a data set that contains the key features.
- The "constructed" data set " DATA FOR FIXED EFFECTS.dta" is found on Moodle.
- The data include 2000 individuals.
- Each individual in this file is observed 30 times, that is 30 "years", which means that the data contains 60,000 person-year observations.
- The main variables include, wage in the current year, age, potential experience (square), schooling, person fixed effect and two binary indicators:
  - whether person *i* ever worked in the financial sector and
  - whether she works in this sector in the current year.
- The *causal impact* of finance on wages was constructed to be 0.5. Below is the list of variables.

The Goal

- The casual model exhibits the following form:

$$Y_{it} = \beta_0 + \beta_F F_{it} + \beta_{E1} EXP_{it} + \beta_{E2} EXP_{it}^2 + \beta_T YEAR_{it} + U_{it}. \quad (1)$$

- The parameters that I used to impute wages (Y_{it}) are:

$$Y_{it} = 0 + 0.5 F_{it} + 0.06 EXP_{it} - 0.0015 EXP_{it}^2 + 0.05 YEAR_{it} + U_{it}. \quad (2)$$

- The first year in the data is 1981. The effect of time (year) is 0.05 per year starting in 1982.
- The **goal of our exercise** is to *recover* the *parameter of interest* - **the causal effect of finance on wages** – using the fixed effect statistical model.
- Note that we know the parameters since we constructed the data.
- This allows us to test how well each method works.

Describe the Set of Variables

variable name	storage type	display format	value label	variable label
id	int	%8.0g		ID
YEAR	float	%9.0g		
AGEit	float	%9.0g		Age in year t
Fi	byte	%8.0g		1 if person i worked at some point if finance
Ui	float	%8.0g		Ui
Si	float	%9.0g		Schools years completed
Vit	float	%9.0g		person-specific zero mean shock to Yit
Fit	float	%9.0g		1 if person i works in finance in year t
EXPit	float	%9.0g		Years of potential experience
EXP2it	float	%9.0g		EXP SQ.
Uit	float	%9.0g		Uit=Ui+Vit
Yit	float	%9.0g		wage of perosn i in year t = 0.5*Fit + 0.05*Time + 0.06*EXP - 0.0015*EXP2 + Uit

More About the Data

- The data is constructed in a twins format.
- For each observation in the first group there is an identical observation in the second with two differences:
  - a one point gap in the fixed effect;
  - only observation in the second group "high fixed effects" will end working in the finance industry.
- The exact timing was randomized.
- So while there is a large gap between the fixed effect i.e. the "ability" of workers in the financial sector and all other sectors, by construction, we should find little (or no) gap between the person specific influence conditional on F_{it}.
- For further details see summary statistics in the next slides.

Summary Statistics

```
. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          id |     60000      1000.5     577.355         1       2000
        YEAR |     60000      1995.5    8.655514      1981       2010
       AGEit |     60000      42.499    8.714582        24         60
          Fi |     60000          .5    .5000042         0          1
          Ui |     60000          .5    .5772669   -.511356   1.485644
-------------+--------------------------------------------------------
          Si |     60000     15.9915    1.032688        12         19
         Vit |     60000    5.61e-10    .9847085   -4.02624   4.285672
         Fit |     60000    .2503833    .4332374         0          1
       EXPit |     60000     20.5075    8.774189         0         40
      EXP2it |     60000    497.5427    366.9087         0       1600
-------------+--------------------------------------------------------
         Uit |     60000          .5    1.141441  -3.968719   5.703316
         Yit |     60000    3.433478    1.293422  -1.571435   8.453669
```

Summary Statistics by Sectors

		All	Fi=0	Fi=1	Fi=1 & Fit==0	Fi=1 & Fit==1
Observations		60000	30000	30000	14977	15023
YEAR		1996	1996	1996	1995	1996
AGEit		42.5	42.5	42.5	42.4	42.6
Ui		0.5	0.0	1.0	1.0	1.0
Si		16.0	16.0	16.0	16.0	16.0
Vit		0.0	0.0	0.0	0.0	0.0
Fit		0.3	0.0	0.5	--	--
EXPit		20.5	20.5	20.6	20.5	20.7
Uit		0.5	0.0	1.0	1.0	1.0
Yit		3.4	2.8	4.1	3.8	4.3

The Relevant STATA Codes

- A few words about the regression procedures in STATA.
- The procedure "**regress**", in short "**reg**", fits a model of *depvar*, that is a left hand side variable, on indepvars, that is right hand side variables, using *linear regression*.
- The "**regress**" code uses OLS. It produces a variety of summary statistics along with the table of regression coefficients.
- There are other regression commands that may be of interest.
- The "**xtreg**" command is one of those. The "**xtreg**" procedure fits regression models to panel data.
- There are several options. To run the fixed-effects option use the option "@, fe i(id)"
- In the next slides you can find the commands summary statistics along with the table of regression coefficients.

Procedure and Outputs: OLS

- The naive regression model with no fixed effects:

```
.  reg Yit Fit YEAR Si EXPit EXP2it

      Source |       SS       df       MS              Number of obs =   60000
-------------+------------------------------           F(  5, 59994) = 4466.94
       Model |  27230.3495     5   5446.0699           Prob > F      =  0.0000
    Residual |  73144.3706 59994  1.21919476           R-squared     =  0.2713
-------------+------------------------------           Adj R-squared =  0.2712
       Total |   100374.72 59999  1.67293988           Root MSE      =  1.1042

------------------------------------------------------------------------------
         Yit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Fit |   1.165641   .0104082   111.99   0.000     1.145241    1.186042
        YEAR |   .0462516   .0044812    10.32   0.000     .0374683    .0550349
          Si |   .0864651   .0061984    13.95   0.000     .0743162     .098614
       EXPit |   .0629844   .0051486    12.23   0.000     .0528931    .0730758
      EXP2it |  -.0014707   .0000631   -23.31   0.000    -.0015944   -.0013471
       _cons |  -91.09604   8.782141   -10.37   0.000    -108.3091   -73.88302
------------------------------------------------------------------------------
```

Differences from Persons' Means

- Generate the mean values of each one of the variables by individual.
  - egen Yim = mean(Yit), by(id);
  - egen Fim = mean(Fit), by(id);
  - egen Sim = mean(Si), by(id);
  - egen EXPim = mean(EXPit), by(id);
  - egen EXP2im = mean(EXP2it), by(id);
  - egen YEARim = mean(YEAR), by(id);
- Generate a new set of variables that is the difference between the value for worker i in year t and the person-specific mean value
  - gen _Yit = Yit - Yim ;
  - gen _Fit = Fit - Fim ;
  - gen _Sit = Si - Sim ;
  - gen _EXPit = EXPit - EXPim ;
  - gen _EXP2it = EXP2it - EXP2im ;
  - gen _YEAR = YEAR - YEARim ;

OLS Using Differences from Persons' Means

- OLS estimates using differences from person-specific means:

```
> reg _Yit _Fit _YEAR _Sit _EXPit _EXP2it ;
note: _Sit omitted because of collinearity
note: _EXPit omitted because of collinearity

      Source |       SS           df       MS      Number of obs   =    60,000
-------------+----------------------------------   F(3, 59996)     =   4606.97
       Model |  13401.6587         3  4467.21955   Prob > F        =    0.0000
    Residual |  58176.0736    59,996  .969665871   R-squared       =    0.1872
-------------+----------------------------------   Adj R-squared   =    0.1872
       Total |  71577.7322    59,999  1.19298209   Root MSE        =    .98472

------------------------------------------------------------------------------
        _Yit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _Fit |   .5007349    .011565    43.30   0.000     .4780675    .5234023
       _YEAR |   .1096825   .0023563    46.55   0.000     .1050643    .1143008
        _Sit |           0  (omitted)
      _EXPit |           0  (omitted)
     _EXP2it |  -.0014767   .0000563   -26.22   0.000    -.0015871   -.0013663
       _cons |   6.67e-09   .0040201     0.00   1.000    -.0078794    .0078794
```

Estimate Fixed Effects using STATA Code

- The fixed-effects regression model:

```
. xtreg Yit Fit YEAR Si EXPit EXP2it, fe i(id)
note: Si omitted because of collinearity
note: EXPit omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =     60000
Group variable: id                              Number of groups   =      2000

R-sq:  within  = 0.1872                         Obs per group: min =        30
       between = 0.6121                                        avg =      30.0
       overall = 0.2274                                        max =        30

                                                F(3,57997)         =   4453.47
corr(u_i, Xb)  = 0.2065                         Prob > F           =    0.0000

------------------------------------------------------------------------------
         Yit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Fit |   .5007349   .0117626    42.57   0.000     .4776801    .5237897
        YEAR |   .1096825   .0023965    45.77   0.000     .1049854    .1143797
          Si |           0  (omitted)
       EXPit |           0  (omitted)
      EXP2it |  -.0014767   .0000573   -25.78   0.000     -.001589   -.0013644
       _cons |  -214.8287   4.754294   -45.19   0.000    -224.1471   -205.5103
-------------+----------------------------------------------------------------
     sigma_u |  .58093819
     sigma_e |  1.0015426
```

Fixed Effects using the "areg" Code

- The fixed-effects regression model:

```
. areg Yit Fit YEAR Si EXPit EXP2it, absorb(id)
note: Si omitted because of collinearity
note: EXPit omitted because of collinearity

Linear regression, absorbing indicators                 Number of obs =   60000
                                                        F(  3, 57997) = 4453.47
                                                        Prob > F      =  0.0000
                                                        R-squared     =  0.4204
                                                        Adj R-squared =  0.4004
                                                        Root MSE      =  1.0015

------------------------------------------------------------------------------
         Yit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Fit |   .5007349   .0117626    42.57   0.000     .4776801    .5237897
        YEAR |   .1096825   .0023965    45.77   0.000     .1049854    .1143797
          Si |  (omitted)
       EXPit |  (omitted)
      EXP2it |  -.0014767   .0000573   -25.78   0.000     -.001589   -.0013644
       _cons |  -214.8287   4.754294   -45.19   0.000    -224.1471   -205.5103
-------------+----------------------------------------------------------------
          id |    F(1999, 57997) =      7.596   0.000        (2000 categories)
```

Summarizing in One Table

Variables	OLS1 reg	OLS2 reg dif	FE1 xtreg	FE2 areg
Fit	1.166***	0.501***	0.501***	0.501***
	(0.010)	(0.012)	(0.012)	(0.012)
YEAR	0.046***	0.110***	0.110***	0.110***
	(0.004)	(0.002)	(0.002)	(0.002)
Si	0.086***			
	(0.006)			
EXPit	0.063***			
	(0.005)			
EXP2it	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-91.096***	0.000	-214.829***	-214.829***
	(8.782)	(0.004)	(4.754)	(4.754)
Observations	60000	60000	60000	60000
R-square	0.271	0.187	0.187	0.420