Fixed Effects and First Differences in Practice Using the NLSY79

Yona Rubinstein

July 2016

Objective

- In this note we practice the use of the *Fixed Effects* and the 1st - *Difference* methods in **observational data**.
- This practice focuses on the Mincerian's returns to finance.
- Specifically I'll be estimating Mincerian wage regressions using OLS in two ways:
    - controlling for observables such as schooling and person-specific traits;
    - controlling for worker-specific fixed effects.
- I'll control for workers' fixed effects using (i) the FE model and (ii) the 1st-DIF model.
- The data is taken from the National Longitudinal Survey of Youth, the 1979 wave, hereafter **NLSY79**.
- The data set and my codes for this exercise are available on Moodle.

Upload the Data and the Programs

- You can find the (i) data file, (ii) the STATA do file, (iii) the log file and (iv) the results files on my website – yonarubinstein.com.
- Download the zip file titled "FIXED EFFECTS IN PRACTICE USING THE NLSY79" .
- Data is taken from the NLSY79.
- The file contains 21,112 observations, white male full-time salaried workers, aged 25+
- We classified workers into 2 main sectors: (i) finance and (ii) all others.
- Describe the data, that is document the list of variables using the "des" command.
- Summarize the main statistics on the key variables (let's say all variables).
- You can find that in the next slides.

Step 1: Describe the Data

```
              storage  display    value
variable name   type   format     label      variable label
----------------------------------------------------------------------------------------------
id              float  %9.0g                 ID# (1-12686) 79
year            float  %9.0g
_SEXi           byte   %9.0g      vlR0214800
                                             SEX OF R 79
_RACEi          byte   %23.0g     vlR0214700
                                             RACL/ETHNIC COHORT /SCRNR 79
_AFQTi          float  %9.0g                 afqt79
_ILLICTi        float  %9.0g                 standardized illicit score 1980
_ROTTERi        float  %9.0g                 standardized rotter score 1979
_ROSENBERGi     float  %9.0g                 standardized rosenberg score 1979
_Sit            byte   %9.0g                 years of schooling completed
_EXP1it         float  %9.0g                 experience in the labor market
_EXP2it         float  %9.0g                 experience square in the labor market
_FINANCEit      byte   %9.0g                 Finance
_TYPEWRKRit     byte   %26.0g     vTYPEWRKRt
                                             TYPE OF WORKER IN CURRENT YEAR, 25+
_FTFYit         byte   %9.0g                 1 IF WKS>=50 & HRS P WK>=35
_HRSWRKit       int    %9.0g      vlR0215710
                                             # OF HRS WRKD IN P-C YR 79
_WAGERit        float  %9.0g                 TOT INC WAGES AND SALRY CPI ADJUSTED 2010
_HWRit          float  %9.0g                 HOURLY WAGE CPI ADJUSTED 2010
_Wit            float  %9.0g                 log hourly wage cpi adjusted (2010)
```

Step 2: Summary Statistics

Sector	Finance				Others			
Observations	898	898	898	898	20214	20214	20214	20214
	Mean	SD	Min	Max	Mean	SD	Min	Max
_AFQTi	0.76	0.02	0.20	0.99	0.57	0.28	0.01	0.99
_ILLICTi	-0.07	-0.94	0.77	3.88	0.22	1.05	-0.94	4.79
_ROTTERi	-0.37	-1.92	0.89	2.20	-0.16	0.99	-1.92	3.03
_ROSENBERGi	0.32	-2.51	0.94	1.85	0.16	0.97	-2.99	1.85
_Sit	15.67	11.00	1.92	20.00	13.79	2.59	6.00	20.00
_EXP1it	14.20	0.00	8.30	37.00	15.48	8.18	0.00	41.00
_TYPEWRKRit	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
_FTFYit	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
_HRSWRKit	2445	2000	429	6240	2469	516	2000	8736
_WAGERit	93460	3071	77359	333309	59248	44678	70	333309
_HWRit	37.6	1.0	29.7	160.2	24.2	17.3	0.0	160.2
_Wit	3.41	0.02	0.64	5.08	3.01	0.59	-3.64	5.08

Step 3: Test Balanced Treatment Comparison Groups

- Phillipon's and Reshef's (2012) empirical strategy relies (implicitly) on wage comparisons between workers in the financial sector and workers in all other sectors controlling for **observables**.
- Workers in finance are the "treatment" group whereas all others are the "comparison" group.
- To check whether this particular econometric setting makes sense we begin by testing whether these groups, that is finance and all others, mimic a randomized trial on relevant observed variables, other than wages.
- Specifically, let's calculate the means of key variables by group and test whether the unconditional means are different.
- The next slides reports average values of (i) cognitive, (ii) non-cognitive traits and (iii) schooling for these groups.
- There are 4 columns. The first 2 report average values. The next two report differences and p-values.

	Finance	Others	Dif.	P-Val
Observations	898	20214		
Early Determined Skills				
_AFQTi	0.76	0.57	**0.19**	0.000
_ROSENBERGi	0.32	0.16	**0.16**	0.101
_ROTTERi	-0.37	-0.16	**-0.20**	0.030
_ILLICTi	-0.07	0.22	**-0.30**	0.000
Schooling and Experience				
_Sit	15.7	13.8	**1.9**	0.000
_EXP1it	14.2	15.5	**-1.3**	0.005
Labor Market Outcomes				
_HRSWRKit	2445	2469	**-24**	0.465
_WAGEit	71472	42949	**28523**	0.000
_HWRit	38	24	**13**	0.000
_Wit	3.41	3.01	**0.40**	0.000

Step 3: The Causal Model

- Following Mincer (1974) let's assume that the *causal model* of log hourly wages exhibits the following functional form:

$$
\begin{aligned}
Y_{it} \quad = \quad & \beta_0 + \beta_F F_{it} + \beta_C C_i + \beta_{NC} NC_i + \beta_S S_{it} + \beta_{E1} EXP_{it} + \beta_{E2} EXP_{it}^2 \\
& + \delta_{81} YR81 ... + \delta_{12} YR12 + U_{it},
\end{aligned}
$$

- Where:

$$U_{it} = \theta_i + \varepsilon_{it}. \qquad (1)$$

  - Y_{it} = log hourly wage of worker i in year t.
  - F_{it} = a binary variable that equals 1 if worker i is employed in the financial sector in year t and 0 otherwise.
  - C_i; NC_i = cognitive and non-cognitive traits measured early in life.
  - S_{it} = years of schooling completed by worker i in year t.
  - EXP_{it} = years of potential labor market experience (worker i; year t).
  - X_{it} = a vector of other observables that might influence wages.
  - $YR81 ... YR12$ = a set of dummy variables that equal to 1 when $YEAR = t$ so in 1981 $YR81 = 1$ and all other $YR@ = 0$.

Step 4: Estimating Fixed Effects Models

- There are basically 3 ways to run/estimate that the fixed effects model in equation (**??**) using STATA:

1. Add manually dummy variables for each subject: by that we generate differences from each person means;
2. Use the command <u>*areg*</u>; "**areg** Y_{it} F_{it} C_i NC_i S_{it} EXP_{it} $EXP2_{it}$ $YR81 \ldots YR12$, absorb (id)". This is equivalent to adding a dummy for each subject.
3. Use the command <u>*xtreg*</u>; "**xtreg** Y_{it} F_{it} C_i NC_i S_{it} EXP_{it} $EXP2_{it}$ $YR81 \ldots YR12$, fe i (id)"

- The next slides report estimates of the model above using our data.
- The explanatory variables include, in addition to and indicator whether person i works in the financial sector in time t: (i) experience; (ii) education; (iii) personality traits, (iv) time effects.

```
. xi: reg _Wit _FINANCEit i.year  ;
i.year           _Iyear_1982-2012    (naturally coded; _Iyear_1982 omitted)

      Source |       SS           df       MS      Number of obs   =    21,112
-------------+----------------------------------   F(22, 21089)    =    113.06
       Model |  799.589425        22  36.3449739   Prob > F        =    0.0000
    Residual |  6779.29905    21,089  .32146138    R-squared       =    0.1055
-------------+----------------------------------   Adj R-squared   =    0.1046
       Total |  7578.88847    21,111  .35900187    Root MSE        =    .56698

------------------------------------------------------------------------------
        _Wit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _FINANCEit |   .3816138   .0193421    19.73   0.000     .3437017    .4195259
       _cons |   2.755617   .0459906    59.92   0.000     2.665472    2.845762
------------------------------------------------------------------------------

. estimates store OLS1, title((CRUDE1));
```

```
. xi:reg _Wit _FINANCEit _AFQTi _ROSENBERGi _ROTTERi _Sit _EXP1it _EXP2it i.year ;
i.year            _Iyear_1982-2012    (naturally coded; _Iyear_1982 omitted)

      Source |       SS           df       MS      Number of obs   =    21,112
-------------+----------------------------------   F(28, 21083)    =    309.11
       Model |  2205.77599        28  78.777714    Prob > F        =    0.0000
    Residual |  5373.11248    21,083  .254855214   R-squared       =    0.2910
-------------+----------------------------------   Adj R-squared   =    0.2901
       Total |  7578.88847    21,111  .35900187    Root MSE        =    .50483

------------------------------------------------------------------------------
        _Wit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _FINANCEit |   .1931051   .0174483    11.07   0.000     .1589051     .227305
      _AFQTi |   .4488653   .0169762    26.44   0.000     .4155908    .4821399
 _ROSENBERGi |   .0246251   .0039242     6.28   0.000     .0169334    .0323168
    _ROTTERi |  -.0317962   .0038052    -8.36   0.000    -.0392547   -.0243377
        _Sit |   .0614077   .0026901    22.83   0.000      .056135    .0666804
     _EXP1it |   .0578167   .0029223    19.78   0.000     .0520887    .0635447
     _EXP2it |  -.0014738   .0000683   -21.58   0.000    -.0016076     -.00134
       _cons |   1.354071   .0583719    23.20   0.000     1.239658    1.468485
------------------------------------------------------------------------------

. estimates store OLS2, title((RESID1));
```

```
. xi:areg _Wit _FINANCEit _AFQTi _ROSENBERGi _ROTTERi _Sit _EXP1it _EXP2it i.year, absorb(id);
i.year            _Iyear_1982-2012    (naturally coded; _Iyear_1982 omitted)

Linear regression, absorbing indicators         Number of obs     =     21,112
                                                F(  24,  18578)   =     255.41
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7294
                                                Adj R-squared     =     0.6925
                                                Root MSE          =     0.3323

------------------------------------------------------------------------------
        _Wit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _FINANCEit |    .076569   .0220506     3.47   0.001     .0333478    .1197902
      _AFQTi |          0  (omitted)
 _ROSENBERGi |          0  (omitted)
    _ROTTERi |          0  (omitted)
        _Sit |          0  (omitted)
     _EXP1it |   .0357336   .0948306     0.38   0.706     -.150143    .2216102
     _EXP2it |  -.0016691   .0000509   -32.79   0.000    -.0017689   -.0015693
 _Iyear_1983 |   .0023816      .1017     0.02   0.981    -.1969598    .2017229
       _cons |   2.471255   .1756713    14.07   0.000     2.126923    2.815587
-------------+----------------------------------------------------------------
          id |   F(2509, 18578) =     15.522   0.000        (2510 categories)
```

```
. xi:xtreg _Wit _FINANCEit _AFQTi _ROSENBERGi _ROTTERi _Sit _EXP1it _EXP2it i.year, fe i(id);
i.year            _Iyear_1982-2012    (naturally coded; _Iyear_1982 omitted)

Fixed-effects (within) regression               Number of obs      =     21,112
Group variable: id                              Number of groups   =      2,510

R-sq:                                           Obs per group:
     within  = 0.2481                                         min =          1
     between = 0.0586                                         avg =        8.4
     overall = 0.1320                                         max =         22

                                                F(24,18578)        =     255.41
corr(u_i, Xb)  = -0.0295                        Prob > F           =     0.0000

------------------------------------------------------------------------------
         _Wit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _FINANCEit |    .076569   .0220506     3.47   0.001     .0333478    .1197902
       _AFQTi |          0  (omitted)
  _ROSENBERGi |          0  (omitted)
     _ROTTERi |          0  (omitted)
         _Sit |          0  (omitted)
      _EXP1it |   .0357336   .0948306     0.38   0.706     -.150143    .2216102
      _EXP2it |  -.0016691   .0000509   -32.79   0.000    -.0017689   -.0015693
        _cons |   2.471255   .1756713    14.07   0.000     2.126923    2.815587
-------------+----------------------------------------------------------------
      sigma_u |    .527535
      sigma_e |  .33227082
          rho |  .71596416   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(2509, 18578) = 15.52                Prob > F = 0.0000
```

Regression Coefficients W/O Fixed Effects

	OLS	OLS |X	areg	xtreg
_FINANCEit	**0.382*****	**0.193*****	**0.077*****	**0.077*****
	(0.019)	(0.017)	(0.022)	(0.022)
_AFQTi		0.449***	0.000	0.000
		(0.017)		
_ROSENBERGi		0.025***		
		(0.004)		
_ROTTERi		-0.032***		
		(0.004)		
_Sit		0.061***		
		(0.003)		
_EXP1it		0.058***	0.036	0.036
		(0.003)	(0.095)	(0.095)
_EXP2it		-0.001***	-0.002***	-0.002***
		(0.000)	(0.000)	(0.000)
Constant	2.756***	1.354***	2.471***	2.471***
	(0.046)	(0.058)	(0.176)	(0.176)
Observations	21112	21112	21112	21112
R-square	0.106	0.291	0.729	0.248

Regression Coefficients of Wages on Finance W/O Fixed Effects

- The regression coefficient on finance drops as we add controllers.
- Adding schooling and measures of cognitive and non-cognitive skills shrinks the coefficient on finance.
- Controlling for person fixed effects reduces the regression coefficient by more than half.
- We obtain identical estimates when we use deviations from each person means – using the command **areg** – and by using the fixed effects command **xtreg**.
- It shouldn't surprise us since the fixed effects model is deviations from subjects' means.
- Next we estimate the finance wage premium before and after bank deregulation.

Step 5: The Effect of Finance Before and After 1996

- Is the wage premium on employment in the financial sector after bank deregulation higher then during the years before (prior to 1996)? Does the causal impact of finance on log hourly wage increased after bank deregulation?
- There are different ways to address this question.
- We estimate the log hourly wage model twice: (i) for the years prior to 1996; (ii) for the years since 1996:

$$Y_{Tit} = \beta_{T0} + \beta_{TF} F_{it} + \beta_{TS} S_{it} + \beta_{TE1} EXP_{it} + \beta_{TE2} EXP_{it}^2 + \beta_{TX} X_{it} + U_{T} \tag{2}$$

- Where $T = 0$ for the years prior to 1996 and $T = 1$ for the years since 1996.This means that we estimated the model twice: (i) using the years prior to 1996 and (ii) using the years since 1996.
- We estimate the model with and without fixed effects. We report our findings in the next slide.
- What do you learn from the findings?

Regression Coefficients of Wages on Finance Before and After 1996

Variables	Year <1996		Year >=1996	
	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)
_FINANCEit	**0.158*** **	**0.090*** **	**0.304*** **	**0.036**
	(0.022)	(0.030)	(0.029)	(0.036)
_Sit	0.108***		0.121***	
	(0.003)		(0.004)	
_EXP1it	0.091***	0.186**	0.059***	0.084***
	(0.006)	(0.081)	(0.008)	(0.010)
_EXP2it	-0.002***	-0.003***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.889***	2.119***	0.671***	2.011***
	(0.066)	(0.185)	(0.113)	(0.256)
Observations	11770	11770	9342	9342
R-square	0.163	0.777	0.240	0.798

Step 6:Fixed Effects in Wage Growth

- So far we assumed that the following *causal model* of log hourly wages:

$$Y_{it} = \beta_0 + \beta_F F_{it} + \beta_C C_i + \beta_{NC} NC_i + \beta_S S_{it} + \beta_{E1} EXP_{it} + \beta_{E2} EXP_{it}^2 + \delta \tag{3}$$

- Where:

$$U_{it} = \theta_i + \varepsilon_{it}. \tag{4}$$

- What if people differ not only on their level of hourly wages, that is on their fixed effect θ_i, but they also differ on the rate at which their wages grow from year to year?
- We can allow the following:

$$U_{it} = \theta_i + \lambda_i \cdot t + \varepsilon_{it}. \tag{5}$$

- Can we deal with that?
- Yes we can!!!

Step 6: Robustness, Fixed Effects vs Differences

- The first difference equation – the DIF model – takes the following form:

$$\Delta Y_{it} = \lambda_0 + \beta_F \Delta F_{it} + \beta_{E1} \Delta EXP_{it} + \beta_{E2} \Delta EXP_{it}^2 + \delta_{81} YR81 ... + \delta_{12} YR1 \quad (6)$$

- Where $\Delta F_{it} = F_{it} - F_{it-1}$, $\Delta EXP_{it} = EXP_{it} - EXP_{it-1}$, $\Delta EXP_{it}^2 = EXP_{it}^2 - EXP_{it-1}^2$ and:

$$V_{it} = \lambda_i + (\varepsilon_{it} - \varepsilon_{it-1}) \,. \quad (7)$$

- Note that while the model above is less robust to measurement errors in the year workers switch between sectors, it allows to control for person fixed effects in wage profiles by using fixed effects model to estimate the equation above.
- The next slides reports estimates of fixed effects model on (i) log hourly wages (ii) OLS on DIF and (iii) FE on DIF

Fixed Effects, Differences and Differences with Fixed Effects

	Levels (1)	Levels (2)	1st-Dif (3)	1st-Dif (4)
_FINANCEit	**0.193*** **	**0.077*** **		
	(0.040)	(0.022)		
_ΔFINANCEit			**0.062** **	**0.052** **
			(0.024)	(0.025)
Constant	1.354***	2.471***	0.056	0.029
	(0.107)	(0.176)	(0.037)	(0.019)
Controlling for person fixed effets				
	No	Yes	No	Yes
Observations	21112	21112	18602	18602
R-square	0.291	0.729	0.014	0.092

Take Home Message

- In **observational data** *treatment* and *comparison* groups are often **unbalanced**.
- They differ on **observable** characteristics that might influence the outcomes of interest – in our context log hourly wages.
- Therefore there are good reasons to believe that they also differ on **unobservables** (to the econometrician) characteristics.
- Fixed Effects models are a useful tool to **account** for *selection* on **time invariant** *unobserved factors*.
- They are sensitive to **measurement errors** in *treatment* status and its *timing*.
- We can control for person-specific trends in outcomes by estimating DIF model with fixed effects.
- Note that we should ask ourselves (i) why people switch sectors and (ii) whether the timing is indeed exogenous to changes in their productivity or other factors affecting their pay.