HANDOUT 15 – FIXED EFFECTS

AGENDA

- Introduction: Is richer people happy? (Part I)
- Examples: Family Fixed Effects (Part II)
- Takeaways
- Vocabulary

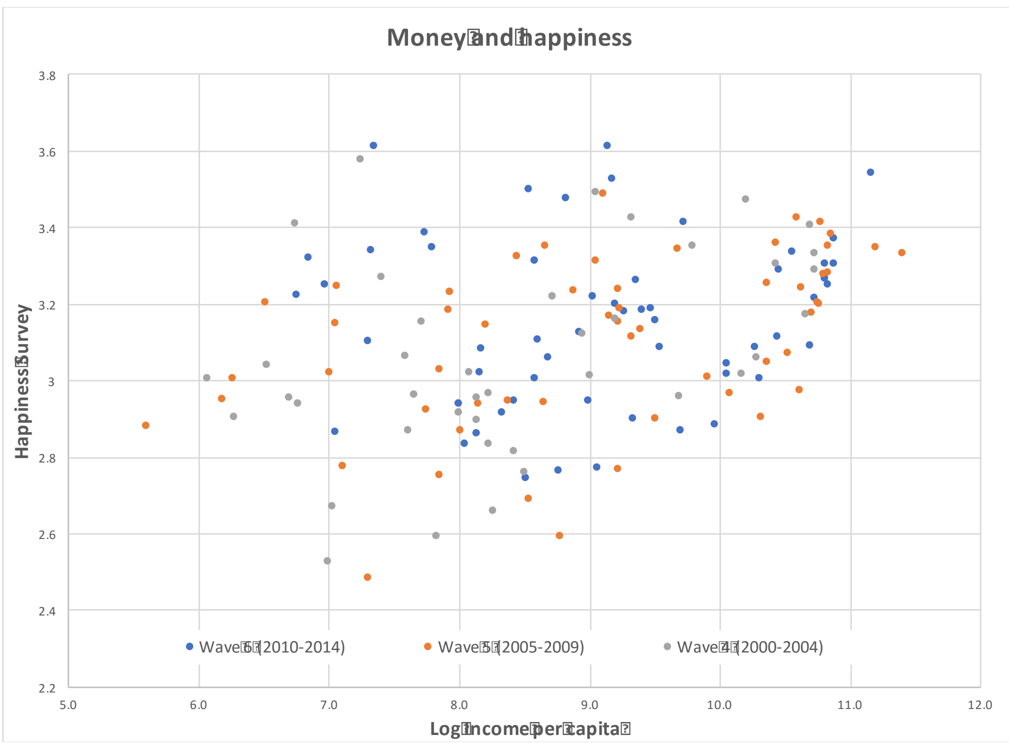
BIBLIOGRAPHY FOR TODAY´S CLASS

- Deming (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. (*) (Focus on the abstract, introduction and section 3B.)
- Stock and Watson (2007), 10.3, 10.4 (**)

INTRODUCTION

- We now know two strategies for dealing with omitted variable bias when the omitted variable cannot be measured (unobservable):
  - Randomized experiments – Researcher creates counterfactual.
  - Instrumental variables – Nature/policy creates counterfactual.
- Fixed effects regressions are another way:
  - Fixed effects – Counterfactual exists within some group.

Money (income per capita, WDI) and happiness (World Happiness Survey)



Money and happiness: This is how the data looks like

WAVE 4

country	hi	gdp	lgdp
Albania	2.59013081	2501.73	7.82
Algeria	2.96443009	3736.41	8.23
Argentin	3.12047243	7721.06	8.95
Banglade	2.90253663	530.68	6.27
Bosnia	3.02016807	3221.56	8.08
Canada	3.40673566	44292.31	10.70
Chile	3.15926242	9830.91	9.19
China	2.86847401	2020.09	7.61
India	2.95325208	808.27	6.69
Indonesi	3.15276384	2235.56	7.71
Iran	2.81295586	4542.36	8.42
Iraq	2.6569438	3862.44	8.26
Israel	3.01776648	26203.28	10.17
Japan	3.17183948	42352.20	10.65
Jordan	2.9146142	2970.67	8.00
South Ko	2.95583344	16088.67	9.69
Kyrgyzst	3.03961349	685.82	6.53
Mexico	3.49047923	8496.73	9.05
Moldova	2.52706838	1087.78	6.99
Morocco	2.96317053	2115.74	7.66
Nigeria	3.57764578	1410.64	7.25
Pakistan	2.93819666	864.31	6.76
Peru	2.95460606	3409.74	8.13
Philippi	3.26711178	1650.03	7.41
Puerto R	3.47214484	26881.46	10.20
Saudi Ar	3.35223484	17822.78	9.79
Singapor	3.3032515	33895.52	10.43

Cross-sectional data

. . .

WAVE 5

country	hi	gdp	lgdp
Andorra	3.20260787	46533.29	10.75
Argentin	3.16733861	9403.79	9.15
Australi	3.28025484	50502.71	10.83
Brazil	3.23863626	10140.11	9.22
Bulgaria	2.59301138	6450.78	8.77
Canada	3.41168284	47764.73	10.77
Chile	3.13426852	12045.20	9.40
China	2.93629932	3448.48	8.15
Colombia	3.34968519	5789.10	8.66
Cyprus	3.25357485	31618.74	10.36
Ethiopia	2.88152599	272.29	5.61
Finland	3.19940782	47206.71	10.76
France	3.24248505	40932.55	10.62
Georgia	2.75270271	2569.28	7.85
Germany	2.97348666	40741.85	10.62
Ghana	3.24527073	1166.95	7.06
Guatemal	3.23123121	2768.45	7.93
Hong Kon	2.90441775	30096.34	10.31
Hungary	2.90019965	13392.18	9.50
India	3.01854634	1108.10	7.01
Indonesi	3.18410468	2747.80	7.92
Iran	2.94267273	5689.79	8.65
Italy	3.07057643	37259.33	10.53
Japan	3.17729831	44593.39	10.71
Jordan	3.14428687	3644.40	8.20
South Ko	3.00916672	20005.50	9.90
Malaysia	3.31057453	8560.55	9.05

Cross-sectional data

. . .

WAVE 6

country	hi	gdp	lgdp
Algeria	2.94	4560.97	8.43
Azerbaij	3.06	5927.03	8.69
Argentin	3.18	10527.31	9.26
Australi	3.30	53157.46	10.88
Bahrain	2.88	21269.68	9.97
Armenia	3.08	3546.89	8.17
Brazil	3.26	11646.62	9.36
Belarus	2.76	6424.58	8.77
Chile	3.08	13948.22	9.54
China	3.01	5339.61	8.58
Colombia	3.48	6795.06	8.82
Cyprus	3.09	28755.51	10.27
Ecuador	3.50	5098.32	8.54
Estonia	2.87	16232.03	9.69
Georgia	2.86	3428.00	8.14
Germany	3.09	43879.95	10.69
Ghana	3.34	1525.00	7.33
Hong Kon	3.11	34224.30	10.44
India	3.10	1486.81	7.30
Iraq	2.74	4983.70	8.51
Japan	3.22	45407.72	10.72
Kazakhst	3.20	9922.44	9.20
Jordan	3.02	3497.91	8.16
South Ko	3.04	23188.88	10.05
Kuwait	3.33	38413.39	10.56
Kyrgyzst	3.32	938.83	6.84
Lebanon	2.95	8021.70	8.99

Cross-sectional data

. . .

- To control for events occurring in different waves that might impact HI:
    - We have learned that in order to allow for different intercepts in the case of each wave we should: _______________________________________.

$$HI_{iw} = \beta_0 + \beta_1 lGDPP_{iw} + \lambda_1 Wave5 + \lambda_2 Wave6 + \varepsilon_{iw}$$

```
. reg hi lgdp D5 D6

------------------------------------------------------------------------------
          hi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lgdp |   .0625514   .0144357     4.33   0.000     .0340247     .091078
          D5 |   -.001199   .0125092    -0.10   0.924    -.0259187    .0235208
          D6 |  -.0081951   .0124199    -0.66   0.510    -.0327383     .016348
       _cons |   2.581334   .1648522    15.66   0.000     2.255565    2.907102
------------------------------------------------------------------------------
```

    - What can we conclude from the previous regression?

- What would have happened if there were 30 or 50 waves?

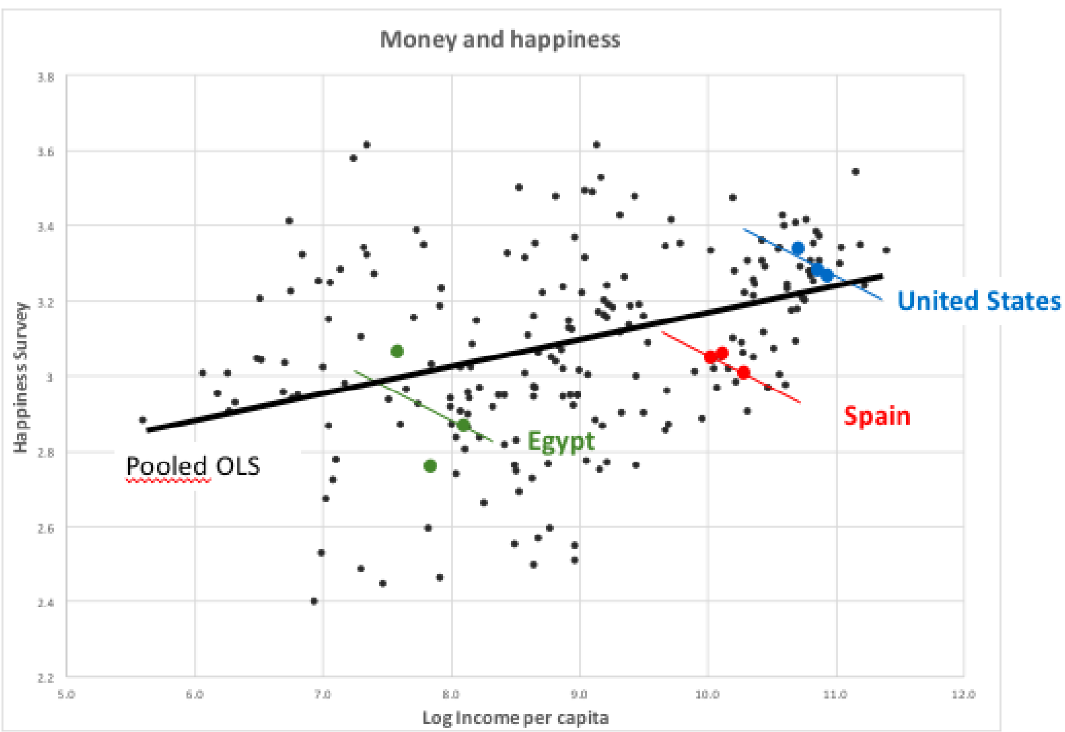
- How would we interpret β_1 in the equation above?

- Money and happiness: this is how the pooled data looks like – Panel data:

country	hi	gdp	lgdp	wave
Albania	2.59	2,501.73	7.82	4
Algeria	2.96	3,736.41	8.23	4
Algeria	2.94	4,560.97	8.43	6
Andorra	3.20	46,533.29	10.75	5
Argentin	3.12	7,721.06	8.95	4
Argentin	3.17	9,403.79	9.15	5
Argentin	3.18	10,527.31	9.26	6
Armenia	3.08	3,546.89	8.17	6
Australi	3.28	50,502.71	10.83	5
Australi	3.30	53,157.46	10.88	6
Azerbaij	3.06	5,927.03	8.69	6
Bahrain	2.88	21,269.68	9.97	6
Banglade	2.90	530.68	6.27	4
Belarus	2.76	6,424.58	8.77	6
Bosnia	3.02	3,221.56	8.08	4
Brazil	3.24	10,140.11	9.22	5
Brazil	3.26	11,646.62	9.36	6
Bulgaria	2.59	6,450.78	8.77	5
Burkina	3.01	526.15	6.27	5
Canada	3.41	44,292.31	10.70	4
Canada	3.41	47,764.73	10.77	5
Chile	3.16	9,830.91	9.19	4
Chile	3.13	12,045.20	9.40	5
Chile	3.08	13,948.22	9.54	6
China	2.87	2,020.09	7.61	4
China	2.94	3,448.48	8.15	5
China	3.01	5,339.61	8.58	6
Colombia	3.35	5,789.10	8.66	5
Colombia	3.48	6,795.06	8.82	6

- I can arrange all observations in a single panel, by adding a column that specifies to which wave does the observation belong.
- In that way we can see that we have various observations (waves) per country in time.

Money (Income per capita logs, WDI) and happiness (World Happiness Survey)



- What do you think is going on?

- Is richer people happier?

Fixed Effects: Measuring the impact of money on happiness **<u>within</u>** countries

- If we suspect that different factors (observable or not) within countries are correlated with money and happiness, or impact in some way with the relationship money affects happiness (i.e. culture, values).

$$HI_{iw} = \beta_0 + \beta_1 lGDPPC_{iw} + \lambda_1 Country1 + \lambda_2 Country2 + \cdots + \lambda_{n-1} Country(n-1) + \varepsilon_{iw}$$

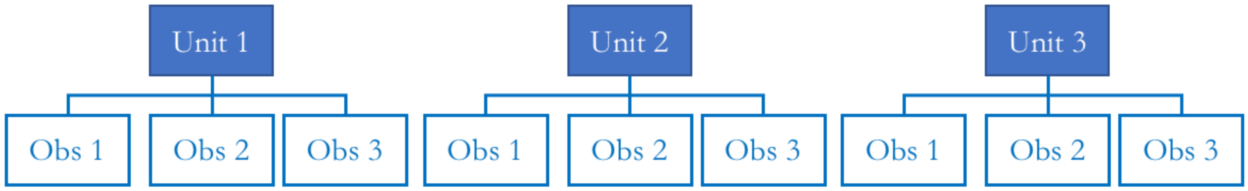
- To make it simpler, we will express the specification above as:

$$HI_{iw} = \beta_0 + \beta_1 lGDPPC_{iw} + \lambda_i + \varepsilon_{iw}$$

- All dummy variables gathered in λ_i are called "country fixed effects".
- The slope coefficient β_1 tells us the **average within-group change** in Y associated with a one-unit **within-group** change in X.
- Whatever sentence you write interpreting β_1 must make clear that the comparison is being done within the group (not across groups).

Fixed Effects: An introduction

- Fixed effects are relevant when I have different observations per unit, but my variable of interest varies within unit:



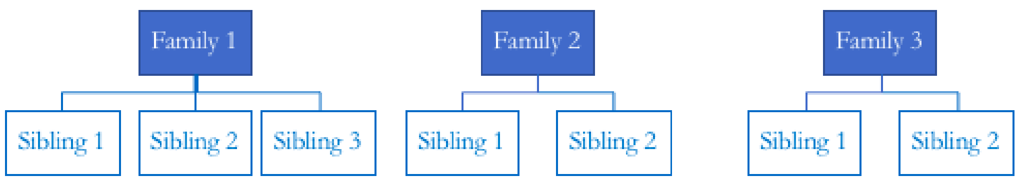
- The idea is to exploit the fact that observations are in different groups, to control for all those factors observable and unobservable that make groups (units) different.
- Fixed effects are a form of multivariate regression with lots of dummies.
- Conceptually – the difference is that we are doing this because we might:
    - Not know what needs to be controlled for.
    - Not able to measure well what needs to be controlled for.
- **We might not know precisely what it is about each group that might be generating OVB** – but we control for all at the unit level.
- Fixed effects **will control for all factors that differ across groups** but **not factors that differ across individuals within groups**.

EXAMPLE: FAMILY FIXED EFFECTS

- The U.S. government funds a pre-school program for low income children called Head Start.
- Policy question: Does Head Start improve kids' educational outcomes?
- *David Deming* – HKS graduate and professor – explored this.
- Why can we not simply compare the outcomes of kids who attend Head Start and those who don't?

- Data: National Longitudinal Survey of Youth's Children and Young Adult cohort.
- Since 1986 this data has tracked all the children from a number of families.

Family ID	Sibling	Head Start	Cognitive test Z-score
1	1	1	0.42
1	2	1	-0.03
1	3	1	-0.19
2	1	0	0.21
2	2	0	0.05
...			
N	1	1	0.09
N	2	0	-0.11



- Empirical challenge is that families who send kids to Head Start may differ in important ways from families who don't.
- *Deming* estimated family fixed effects regressions by including a dummy variable for all (but one) of the families:

$$TestScore_{if} = \beta_0 + \beta_1 HeadStart_{if} + \lambda_1 Family1 + \lambda_2 Family2 + \cdots + \varepsilon_{if}$$

  TestScore = cognitive test Z-score of sibling *i* from family *f*.

  HeadStart = 1 if sibling *i* from family *f* attended Head Start.

  Family2 = 1 if sibling *i* from family *f* is in family 2.

- More typically, the fixed effects are written as a single symbol:

$$TestScore_{if} = \beta_0 + \beta_1 HeadStart_{if} + \lambda_f + \varepsilon_{if}$$

- The family fixed effects control for differences between families that do not vary across sibling.
- As a result, family fixed effects exploit variation within families (sometimes called the **within estimator**).
- The resulting estimates are driven by families that send _______________ of their children to Head Start.
- Families that send ___________________ of their children to Head Start do not affect the estimate.
- Below are *Deming*'s main estimates of the impact of Head start on children's test scores at ages 7-10.
- Columns 1-3: OLS with increasingly rich controls.
- Columns 4-5: Family fixed effects regression.

Table 3—The Effect of Head Start on Cognitive Test Scores

	(1)	(2)	(3)	(4)	(5)
Head Start					
Ages 7–10	–0.116	0.040	0.067	0.116*	0.133**
	(0.072)	(0.065)	(0.061)	(0.060)	(0.060)
Pre-treatment covariates	N	Y	Y	N	Y
Sibling fixed effects	N	N	N	Y	Y

[Source: Deming, David. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start," American Economic Journal: Applied Economics 2009, 111–34.]

- How would you interpret the coefficient in column 1?

- How would you interpret the coefficient in column 4?

- This study claims that Head Start raises kids' age 7-10 test scores.
- What threats to internal validity do the fixed effects take care of?

- What threats to the interval validity of the study are not eliminated by including family fixed effects?

- We can use causal language if we are convinced that the fixed effects have eliminated the main threats to internal validity!

TAKEAWAYS

- Fixed effects allow us to control for some types of confounding variables even when those variables are unobservable.
- The key is to have multiple observations within given units.
- Fixed effects coefficients measure the relationship between within-group differences in treatment status with within-group differences in outcomes.
- If you believe that the most important sources of omitted variable bias are eliminated by fixed effects, you can then use causal language to interpret your regression coefficients.
- Be aware, however, that there could still be other unobserved factors that could explain your results.
- The question is how important those other factors are.

VOCABULARY

- Cross-sectional data: a data set collected by sampling a population at a given point in time.
- Panel data: a data set constructed from repeated cross sections over time. With a balanced panel, the same units appear in each time period. With an unbalanced panel, some units do not appear in each time period, often due to attrition.
- Pooled OLS: OLS estimation with independently pooled cross sections, panel data, or cluster samples, where the observations are pooled across time (or group) as well as across the cross-sectional units.
- Fixed Effects: statistical model in which the model parameters are fixed or non-random quantities. This is in contrast to random effects models and mixed models in which all or some of the model parameters are considered as random variables.
- Within Estimator: for the fixed effects model, the estimator obtained by applying pooled OLS to a time-demeaned equation. It is used to refer to an estimator for the coefficients in the regression model including those fixed effects.

A LAST THOUGHT…

- "…the estimators used to control for fixed effects typically remove both good and bad variation. In other words, these transformations may kill some of the omitted variables bias bathwater, but they also remove much of the useful information in the baby, the variable of interest…

Mostly harmless econometrics, Angrist & Pischke (pp. 226)