Lab L

Kate Brandt

March 5, 2019

```
# Setting up and loading data
library(plm)
library(foreign)
library(readstata13)
library(dplyr)
library(tidyverse)
library(haven)
library(olsrr)
occhist08 <- read_dta("./data/occhist08.dta")
```

Q1.

a. What is the age range of NLSY respondents in 1982?

Answer: 17 - 25

```
occ1982 <- occhist08[occhist08$year==1982,]
summary(occ1982$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   17.00   19.00   21.00   20.85   23.00   25.00       8
```

b. What is the average value of occfem (i.e. % occupation female) for men and women?

Answer: males = 27.8 %, females = 67.2 %

```
occhist08$sex <- factor(occhist08$sex,
                            levels = c(1,2),
                            labels = c("male", "female"))

    occhist08 %>%
        group_by(sex) %>%
        summarize(pctfem = mean(occ_fem, na.rm=TRUE), n = n())
```

```
## # A tibble: 3 x 3
##   sex    pctfem     n
##   <fct>   <dbl> <int>
## 1 male    0.278 79924
## 2 female  0.672 73114
## 3 <NA>    0.221     8
```

Q2. Find 3 interesting career histories

(A) Career of an artist: The search originally sought to follow an author. This case shows a woman who bounced between creative jobs (195 = editors and reporters, 194 = artists, performers, and related workers). It's interesting to notice that there is no linear progression here. She jumps back and forth between these three occupation titles, having only been an author one year. She started off her career as a reporter/editor (these are rather broad to be grouped into one code I would argue), and seemed to switch career paths when her wages decreased in 1984.

```
# Author - 183
 author <- occhist08 %>%
    arrange(caseid, year)%>%
    group_by(caseid) %>%
    filter(any (occ2==183)) %>%
    select(caseid, year, sex, occ, occ2, wage_re, hgc, job)
  table(author$caseid)
```

```
##
##   367   904  1291  1503  2331  2515  2613  2684  2740  2901  3016  3172
##    18    11    13    16    18    14    19    17    15    14    18    14
##  3251  3806  3808  3843  3946  4116  4287  4373  5675  6485  7527  8951
##    18    14    14     6    16    19    13    17    18    17    15    14
##  9610  9744 11924
##    12    12    19
```

```
# Check out case 367
  author[author$caseid==367,]
```

```
## # A tibble: 18 x 8
## # Groups:   caseid [1]
##    caseid  year sex                          occ  occ2 wage_re   hgc   job
##     <dbl> <dbl> <fct>                    <dbl+lbl> <dbl>   <dbl> <dbl> <dbl>
##  1    367  1980 female 276 [Cashiers        ]        NA   11.5     13     1
##  2    367  1981 female 195 [Editors and reporter~    NA    2.76    14     1
##  3    367  1982 female 195 [Editors and reporter~   195    7.83    15     1
##  4    367  1983 female 195 [Editors and reporter~   195   11.4     16     1
##  5    367  1984 female 195 [Editors and reporter~   195    8.86    16     1
##  6    367  1985 female 194 [Artists, performers,~   194   10.3     16     1
##  7    367  1986 female 194 [Artists, performers,~   194   10.8     16     1
##  8    367  1987 female 195 [Editors and reporter~   195   12.1     16     1
##  9    367  1988 female 194 [Artists, performers,~   194   16.2     16     1
## 10    367  1989 female 195 [Editors and reporter~   195   16.4     16     1
## 11    367  1990 female 195 [Editors and reporter~   195   15.9     16     1
## 12    367  1991 female 183 [Authors ]               183   15.6     16     1
## 13    367  1992 female 194 [Artists, performers,~   194   17.5     16     1
## 14    367  1993 female 195 [Editors and reporter~   195   15.7     16     1
## 15    367  1994 female 184 [Technical writers   ~   184   16.6     16     1
## 16    367  1996 female 195 [Editors and reporter~   195   16.6     16     1
## 17    367  1998 female 195 [Editors and reporter~    NA   17.8     16     1
## 18    367  2000 female 195 [Editors and reporter~   195   20.8     16     1
```

(B) Trajectory of a hairdresser: This is an interesting case of a woman who seemed to jump from odd-job to odd job in the interest of pursuing higher wages. She started as a hairdresser after 2 years of college (maybe cosmetology school?). Then she jumped from being a supervisor, receptionist, guard, transportation and ticket reservation agent, laborer, and office clerk until she was 28. Ten years later she was working as a manager/administrator. She never went back to school, so it seems her job options were limited to unskilled labor. The gap in the career history would be helpful for understanding what happened in the years between 1990 and 2000.

```
# Hairdresser/cosmetologist - 458
# There are a lot of people who have ever been a hair dresser
# Let's look for someone who was a hairdresser at age 20, and see how their career progressed
 hairdresser <- occhist08 %>%
    arrange(caseid, year)%>%
    group_by(caseid) %>%
```

```
    filter(any (occ2==458 & age == 20)) %>%
    select(caseid, year, sex, occ, occ2, wage_re, hgc, job, age)
  table(hairdresser$caseid)
```

```
##
##   643   942  1037  2560  3006  3128  4080  4469  4663  4922  4992  5103
##    11    19    16     9    12    18    10    17    19    16    12    15
##  5353  8004 10159 11865
##    13    16    16    16
```

```
  hairdresser[hairdresser$caseid==3006,]
```

```
## # A tibble: 12 x 9
## # Groups:   caseid [1]
##    caseid  year sex                    occ  occ2 wage_re   hgc   job   age
##     <dbl> <dbl> <fct>            <dbl+lbl> <dbl>   <dbl> <dbl> <dbl> <dbl>
##  1   3006  1980 female 389 [Administrative~    NA    7.02    12     1    18
##  2   3006  1981 female 389 [Administrative~    NA    7.71    12     1    19
##  3   3006  1982 female 458 [Hairdressers a~   458   94.1     14     1    20
##  4   3006  1983 female 274 [Sales workers,~   274   NA       14     1    21
##  5   3006  1984 female 305 [Supervisors, f~   305    7.95    14     1    22
##  6   3006  1985 female 319 [Receptionists ~   319    9.39    14     1    23
##  7   3006  1986 female 426 [Guards and pol~   426   15.5     14     1    24
##  8   3006  1987 female 318 [Transportation~   318   20.3     14     1    25
##  9   3006  1988 female 889 [Laborers, exce~   889   26.2     14     1    26
## 10   3006  1989 female 318 [Transportation~   318   24.2     14     1    27
## 11   3006  1990 female 379 [General office~   379   21.0     14     1    28
## 12   3006  2000 female  19 [Managers and a~    19   13.7     14     1    39
```

(C) Career of a chemistry teacher: This seems to be the trajectory of a student who put himself through school then applied to medical school and worked odd jobs through both undergrad and med school. The chemistry teacher job appears for one year while it appears he is in medical school (at age 27) where there is also a gap in the wages he earns. He likely worked as a tutor here. He finally becomes at doctor at 31.

```
chemteach <- occhist08 %>%
    arrange(caseid, year)%>%
    group_by(caseid) %>%
    filter(any (occ2==115)) %>%
    select(caseid, year, sex, occ, occ2, wage_re, hgc, job, age)

table(chemteach$caseid)
```

```
##
##  1745  2428  3971  5245  7453 11897
##    15    12    16    17     6    11
```

```
  chemteach[chemteach$caseid==5245,]
```

```
## # A tibble: 17 x 9
## # Groups:   caseid [1]
##    caseid  year sex                    occ  occ2 wage_re   hgc   job   age
##     <dbl> <dbl> <fct>            <dbl+lbl> <dbl>   <dbl> <dbl> <dbl> <dbl>
##  1   5245  1979 male  479 [Farm workers   ~    NA    7.81    10     1    17
##  2   5245  1980 male  479 [Farm workers   ~    NA    7.68    11     1    17
##  3   5245  1981 male  479 [Farm workers   ~    NA   14.7     12     1    18
##  4   5245  1982 male  479 [Farm workers   ~   479    7.34    13     1    20
```

```
##  5   5245  1983 male  479 [Farm workers   ~   479    7.11    14     1    20
##  6   5245  1984 male  813 [Parking lot att~   813    8.18    15     1    22
##  7   5245  1985 male  479 [Farm workers   ~   479   NA       15     1    22
##  8   5245  1986 male  883 [Freight, stock,~   883   12.9     15     1    24
##  9   5245  1987 male   19 [Managers and ad~    19   12.1     15     1    25
## 10   5245  1988 male  637 [Machinists, exc~   633   11.6     17     1    26
## 11   5245  1989 male  115 [Chemistry teach~   115   NA       18     1    27
## 12   5245  1992 male  446 [Health aides, e~   446   10.0     19     1    30
## 13   5245  1993 male   84 [Physicians     ~    84    8.80    20     1    31
## 14   5245  1994 male   84 [Physicians     ~    84   10.9     20     1    32
## 15   5245  1996 male   84 [Physicians     ~    84   60.4     20     1    34
## 16   5245  1998 male   84 [Physicians     ~    NA   59.0     20     1    36
## 17   5245  2000 male   84 [Physicians     ~    84   62.5     20     1    38
```

Q3. Find the career history of someone who makes a successful transition from a heavily female occupation to a predominantly male occupation. Looking at caseid 21 as an example: This woman transitioned from being a secretary (occ_fem = 0.988) in 1985 to being a manager in 1986 (occ_fem = 0.18). She stays in relatively low proportion-female jobs for the rest of her career/the data set.

```
# filter by first criteria: early career in high female occupation
# filter by second criteria: later career in low female occupation
# I set thresholds for high and low as 0.92 and 0.1 respectively- can be changed

transition <- occhist08 %>%
  arrange(caseid, year)%>%
  group_by(caseid) %>%
  filter(any ((occ_fem >= 0.92 & year <= 1985))) %>%
  filter(any ((occ_fem <= 0.1 & year >= 1986))) %>%
  select(caseid, year, sex, occ, occ2, wage_re, hgc, age, occ_fem)
transition[transition$caseid==21,]
```

```
## # A tibble: 19 x 9
## # Groups:   caseid [1]
##    caseid  year sex                  occ  occ2 wage_re   hgc   age occ_fem
##     <dbl> <dbl> <fct>          <dbl+lbl> <dbl>   <dbl> <dbl> <dbl>   <dbl>
##  1     21  1979 female 274 [Sales worker~    NA    7.94    11    17  0.768
##  2     21  1980 female 274 [Sales worker~    NA    7.57    12    18  0.768
##  3     21  1981 female 335 [File clerks ~    NA    8.56    13    19  0.809
##  4     21  1982 female 435 [Waiters and ~   435   10.4     14    20  0.901
##  5     21  1983 female 319 [Receptionist~   319    7.58    15    21  0.961
##  6     21  1984 female 276 [Cashiers    ~   276    9.55    15    22  0.857
##  7     21  1985 female 313 [Secretaries ~   313   11.0     16    23  0.988
##  8     21  1986 female  13 [Managers, ma~    13   10.3     16    24  0.187
##  9     21  1987 female  13 [Managers, ma~    13   13.6     16    25  0.187
## 10     21  1988 female  19 [Managers and~    19   17.2     16    27  0.291
## 11     21  1989 female 256 [Advertising ~   256   13.7     16    28  0.440
## 12     21  1990 female  13 [Managers, ma~    13   21.6     16    29  0.187
## 13     21  1991 female  13 [Managers, ma~    13   17.0     16    30  0.187
## 14     21  1992 female  19 [Managers and~    19   19.7     16    31  0.291
## 15     21  1993 female  19 [Managers and~    19   20.6     16    31  0.291
## 16     21  1994 female 637 [Machinists, ~   633   24.3     16    33  0.0504
## 17     21  1996 female  13 [Managers, ma~    13   25.0     16    34  0.187
```

```
## 18     21  1998 female 166 [Economists  ~    NA   38.3     16    36  0.330
## 19     21  2000 female  13 [Managers, ma~    13   38.3     16    39  0.187
```

Q4. "Genderness" of occupations over time

An interesting story emerges about gendered occupations based on education status, gender, and the year. Over a ten year periods, across male groups, there is a decline in the average proportion of women in their occupation; the opposite trend is observed between women's groups. College educated men and women are in more "gender integrated" occupations, relative to their high-school educated counterparts. This suggests that college education is a tool for promoting equity of gender representation in occupations.

```
occhist_82_92 <- occhist08[occhist08$year==1982 | occhist08$year == 1992,]
occhist_82_92$year <- factor(occhist_82_92$year,
                             levels = c(1982,1992),
                             labels = c("1982", "1992"))
occhist_82_92$educ <- ifelse(occhist_82_92$hgc<=12, 1, 0)
occhist_82_92$educ <- factor(occhist_82_92$educ,
                             levels = c(1,0),
                             labels = c("high school", "college"))

occhist_82_92 %>%
  group_by(sex, year, educ) %>%
        summarize(pctfem = mean(occ_fem, na.rm=TRUE), n = n())
```

```
## # A tibble: 9 x 5
## # Groups:   sex, year [5]
##   sex    year  educ         pctfem     n
##   <fct>  <fct> <fct>         <dbl> <int>
## 1 male   1982  high school   0.289  3666
## 2 male   1982  college       0.411  1180
## 3 male   1992  high school   0.223  2397
## 4 male   1992  college       0.334  1565
## 5 female 1982  high school   0.715  2870
## 6 female 1982  college       0.718  1423
## 7 female 1992  high school   0.644  1822
## 8 female 1992  college       0.635  1741
## 9 <NA>   <NA>  <NA>        NaN         8
```

Q5. Using regression models to estimate effect of proportion of women in an occupant's field.

```
# Pooled OLS for all years
mod.ols <- lm(lnwage ~ sex + occ_fem + hgc + hrswk + tenur + exp802 + pctile_age,
              data=occhist08)

# Fixed effects
mod.fe <- plm(lnwage ~ sex + occ_fem + hgc + hrswk + tenur + exp802 + pctile_age,
             data=occhist08, index=c("caseid", "year"), na.action=na.omit, model="within")
```

```
## at least one couple (id-time) has NA in at least one index dimension in resulting pdata.frame
##  to find out which, use e.g. table(index(your_pdataframe), useNA = "ifany")
```

```
# Random effects
mod.re <- plm(lnwage ~ sex + occ_fem + hgc + hrswk + tenur + exp802 + pctile_age,
                  data=occhist08, index=c("caseid", "year"), na.action=na.omit, model="random")
```

```
## at least one couple (id-time) has NA in at least one index dimension in resulting pdata.frame
##  to find out which, use e.g. table(index(your_pdataframe), useNA = "ifany")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Mar 05, 2019 - 10:27:43 AM

Table 1: Regression Results

	Dependent variable:		
	log wages		
	OLS	*panel linear*	
	OLS	fixed effects	random effects
	(1)	(2)	(3)
sex: female	0.035***		0.035***
	(0.002)		(0.002)
occupation proportion female	−0.125***	−0.141***	−0.130***
	(0.003)	(0.004)	(0.003)
highest grade achieved	0.030***	0.067***	0.032***
	(0.0004)	(0.001)	(0.0004)
hours worked	0.001***	0.0003***	0.001***
	(0.0001)	(0.0001)	(0.0001)
tenure with employer	0.010***	0.016***	0.011***
	(0.0003)	(0.0004)	(0.0003)
accumulated experience	0.001***	0.001***	0.001***
	(0.00001)	(0.00001)	(0.00001)
age percentile	1.239***	1.175***	1.226***
	(0.003)	(0.004)	(0.003)
Constant	1.291***		1.273***
	(0.005)		(0.005)
Observations	144,915	144,915	144,915
R^2	0.679	0.497	0.639
Adjusted R^2	0.679	0.451	0.639
Residual Std. Error	0.302 (df = 144907)		
F Statistic	43,793.070*** (df = 7; 144907)	21,890.840*** (df = 6; 132713)	256,309.300***
Note:			*p<0.1; **p<0.05; ***p<0.01

Interpretation: Between the three models, there are only slight differences between the effect sizes of the coefficients. The OLS and random effects models are extrememly similar, which suggests that they are picking up on the same effects of unobserved factors. Comparing these two models to the fixed effects model, we can see that there are much more noticeable effects of proportion female in the occupation, education, and age percentile. The proportion female is especially interesting to note becauese it picks up the structural effects of a gendered labor force on wages at the individual level. All three models show that there is a negative relationship between proportion female and wages, but the OLS and RE models show that there is a slightly

positive effect for women. Thus, the FE model can show that at the individual level, the participation in a structurally feminized field has a negative effect on one's earnings, regardless of the individual's gender. This is interesting to compare because we know that this effect is not due specifically to the individual's gender within a gendered field; the OLS and RE models can give us this information. The greatest effect on wages seems to be attributed to the individual's relative age within the field. This is interesting, because it seems to operate separately from accumulated experience, which basically has no effect on wages. Considering this, the FE model may not be able to tell the story about how men are able to accumulate more experience because they do not take time for maternity leave, etc. Further analyses would be necessary to disentangle the processes behind gaining working experience.