PS3

Mattar Klein (328780168)

All files can be found on my github at:

https://github.com/MattarKlein/Econometrics-B.git

Question 1

(a)

$$E[D_i|Y_{1i}, Y_{0i}, X_i] = 1 \cdot P(D_i = 1|Y_{1i}, Y_{0i}, X_i) + 0 \cdot P(D_i = 0|Y_{1i}, Y_{0i}, X_i) = P(D_i = 1|Y_{1i}, Y_{0i}, X_i)$$

(b)

$$E[D_i|Y_{1i}, Y_{0i}, p(X_i)] = E[E[D_i|Y_{1i}, Y_{0i}, p(X_i), X_i]] = E[E[D_i|X_i]|Y_{1i}, Y_{0i}, p(X_i)]$$

Where the second equality is b/c of the independence of $\{Y_{1i}, Y_{0i}\}$ and $D_i|X_i$.

(c)

$$E[D_i|X_i] = Pr(D_i = 1|X_i) = p(X_i)$$
$$\rightarrow E[D_i|Y_{1i}, Y_{0i}, p(X_i)] = E[E[D_i|X_i]|Y_{1i}, Y_{0i}, p(X_i)] =$$
$$E[p(X_i)|Y_{1i}, Y_{0i}, p(X_i)] = p(X_i) = E[D_i|X_i]$$

Question 2

(a) Import data

```
. clear all
. use "nswre74.dta"
. append using "cps1re74.dta", gen(obs)
```

Regress (linearly)observable characteristics on treatment in experimental data:

```
. quietly gen age_2 = age^2
. local i = 1
. foreach var in age age_2 ed black hisp married nodeg re74 re75 {
  2. quietly regress `var´ treat if obs == 0, r
  3. eststo a_reg_`i´
  4. local ++i
  5. }
```

Repeat regression

```
. foreach var in age age_2 ed black hisp married nodeg re74 re75 {
  2. quietly regress `var´ treat if obs == 1, r
  3. eststo a_reg_`i´
  4. local ++i
  5. }
```

I will now look at all the observables to see if they are blanced.

Results experimental

```
. forvalues j = 1/3{
  2. local k = 3*(`j´)-2
  3. local r = 3*(`j´)-1
  4. local t = 3*(`j´)
  5. esttab a_reg_`k´ a_reg_`r´ a_reg_`t´, se
  6. }
```

	(1) age	(2) age_2	(3) ed
treat	0.762	40.08	0.257
	(0.684)	(41.38)	(0.179)
_cons	25.05***	677.3***	10.09***
	(0.438)	(26.60)	(0.100)
N	445	445	445

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

	(1) black	(2) hisp	(3) married
treat	0.0163	-0.0482	0.0353
	(0.0356)	(0.0260)	(0.0365)
_cons	0.827***	0.108***	0.154***
	(0.0235)	(0.0193)	(0.0224)
N	445	445	445

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

	(1) nodeg	(2) re74	(3) re75

treat	-0.127**	-11.45	265.1
	(0.0407)	(503.5)	(305.0)
_cons	0.835***	2107.0***	1266.9***
	(0.0231)	(352.9)	(192.5)
N	445	445	445

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

Results observational

```
. forvalues j = 3/6{
  2. local k = 3*(`j´)-2
  3. local r = 3*(`j´)-1
  4. local t = 3*(`j´)
  5. esttab a_reg_`k´ a_reg_`r´ a_reg_`t´, se
  6. }
```

	(1)	(2)	(3)
	nodeg	re74	re75
treat	-0.127**	-11.45	265.1
	(0.0407)	(503.5)	(305.0)
_cons	0.835***	2107.0***	1266.9***
	(0.0231)	(352.9)	(192.5)
N	445	445	445

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

	(1)	(2)	(3)
	age	age_2	ed
treat	-7.409***	-508.5***	-1.682***
	(0.532)	(32.23)	(0.149)
_cons	33.23***	1225.9***	12.03***
	(0.0873)	(6.206)	(0.0227)
N	16177	16177	16177

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

	(1)	(2)	(3)
	black	hisp	married
treat	0.770***	-0.0126	-0.523***
	(0.0268)	(0.0175)	(0.0290)
_cons	0.0735***	0.0720***	0.712***
	(0.00206)	(0.00204)	(0.00358)
N	16177	16177	16177

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

	(1) nodeg	(2) re74	(3) re75
treat	0.412***	-11921.2***	-12118.7***
	(0.0336)	(366.2)	(247.2)
_cons	0.296***	14016.8***	13650.8***
	(0.00361)	(75.68)	(73.31)
N	16177	16177	16177

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

After that long list it seems that the expermental is pretty balanced (after doing so many regressions it makes sense to have some things be statistically significant), however the observational data is not well balanced.

(b) Run regression without and with controls

```
. quietly regress re78 treat if obs == 0, r
. eststo b_reg_1
. quietly regress re78 treat age age_2 ed nodeg married black hisp  if obs == 0, r
. eststo b_reg_2
. quietly regress re78 treat age age_2 ed nodeg married black hisp re74 re75  if obs == 0, r
. eststo b_reg_3

. quietly regress re78 treat if obs == 1, r
. eststo b_reg_4
. quietly regress re78 treat age age_2 ed nodeg married black if obs == 1, r
. eststo b_reg_5
. quietly regress re78 treat age age_2 ed nodeg married black hisp re74 re75 if obs == 1, r
. eststo b_reg_6
```

Results - experimental

```
. esttab b_reg_1 b_reg_2 b_reg_3, se
```

	(1) re78	(2) re78	(3) re78
treat	1794.3**	1670.0*	1675.9*
	(670.8)	(672.5)	(676.9)
age		178.3	141.7
		(252.4)	(255.1)
age_2		-2.084	-1.435
		(4.023)	(4.069)
ed		378.1	385.0*
		(193.5)	(193.8)
nodeg		-106.8	-55.54
		(1050.5)	(1037.9)
married		82.76	-184.9
		(934.2)	(913.8)
black		-2212.5*	-2155.6*
		(996.7)	(1010.9)
hisp		118.6	187.3
		(1356.6)	(1371.6)
re74			0.0815
			(0.107)
re75			0.0508
			(0.124)
_cons	4554.8***	-422.0	-307.0
	(340.2)	(5024.8)	(5044.8)
N	445	445	445

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

.

Results - observational

```
. esttab b_reg_4 b_reg_5 b_reg_6, se
```

	(1) re78	(2) re78	(3) re78
treat	-8497.5***	-3468.2***	793.6
	(581.9)	(611.7)	(618.6)
age		934.3***	-233.7***
		(49.99)	(40.72)
age_2		-12.07***	1.814**
		(0.701)	(0.558)
ed		238.8***	166.8***
		(38.14)	(28.71)
nodeg		-1094.4***	311.8
		(226.5)	(176.4)
married		3205.8***	224.3
		(192.3)	(152.4)
black		-2231.8***	-790.6***
		(267.7)	(197.8)
hisp			-176.0
			(218.3)
re74			0.295***
			(0.0152)
re75			0.471***
			(0.0153)
_cons	14846.7***	-6063.6***	7634.3***
	(76.29)	(920.1)	(737.8)
N	16177	16177	16177

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

The signs in the different datasets are oposite until controlling for all observables which is another sign that the observational data is not well balanced. Even then the effects in the expermiental data seem higher.

(c)

```
. quietly regress treat age age_2 ed nodeg married black re74 re75 if obs == 1, r
. eststo c_reg_1
. quietly predict lm_prop
. quietly logit treat age age_2 ed nodeg married black re74 re75 if obs == 1, vce(robust)
. eststo c_reg_2
. quietly predict log_prop
. quietly probit treat age age_2 ed nodeg married black re74 re75 if obs == 1, vce(robust)
. eststo c_reg_3
. quietly predict pro_prop
```

Results

```
. esttab c_reg_1 c_reg_2 c_reg_3, se
```

	(1) treat	(2) treat	(3) treat
main			
age	0.00390***	0.745***	0.373***
	(0.000738)	(0.0906)	(0.0424)
age_2	-0.0000551***	-0.0121***	-0.00607***
	(0.00000953)	(0.00158)	(0.000720)
ed	0.000805*	-0.0372	-0.0225
	(0.000350)	(0.0386)	(0.0184)
nodeg	0.0174***	1.453***	0.676***
	(0.00302)	(0.282)	(0.138)
married	-0.0182***	-1.442***	-0.672***
	(0.00258)	(0.251)	(0.124)
black	0.107***	3.574***	1.683***
	(0.00825)	(0.213)	(0.0934)
re74	-0.000000295	-0.0000647	-0.0000277
	(0.000000159)	(0.0000392)	(0.0000181)
re75	-0.000000676***	-0.000217***	-0.0000969***
	(0.000000137)	(0.0000434)	(0.0000207)
_cons	-0.0484***	-14.65***	-7.286***
	(0.0121)	(1.370)	(0.656)
N	16177	16177	16177

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

```
. kdensity pro_prop, nograph generate(x fx)
. kdensity pro_prop if obs == 1 & treat == 1, nograph generate(fx0) at (x)
. kdensity pro_prop if obs == 1 & treat == 0, nograph generate(fx1) at (x)
. label var fx0 "Treatment"
. label var fx1 "Control"
. line fx0 fx1 x, sort ytitle(Density)
. graph export dens.pdf, replace
(file dens.pdf written in PDF format)
```

It looks like everything is around 0. I will try to get a better picture by looking at <0.1.

```
. kdensity pro_prop, nograph generate(y gy)
. kdensity pro_prop if obs == 1 & treat == 1 & pro_prop > 0.1, nograph generate(gy0) at (y)
. kdensity pro_prop if obs == 1 & treat == 0 & pro_prop > 0.1, nograph generate(gy1) at (y)
. label var gy0 "Treatment"
. label var gy1 "Control"
. line gy0 gy1 y, sort ytitle(Density >0.1)
. graph export dens2.pdf, replace
(file dens2.pdf written in PDF format)
```

It seems that the control in this case is ditributed very differently from the treatment.

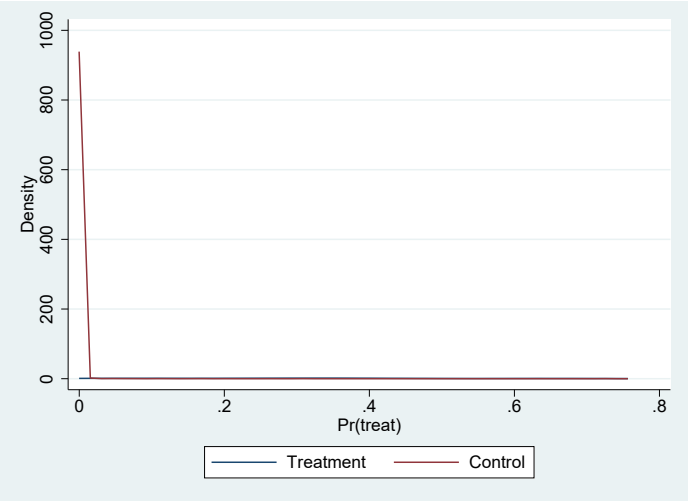

Figure 1: Density

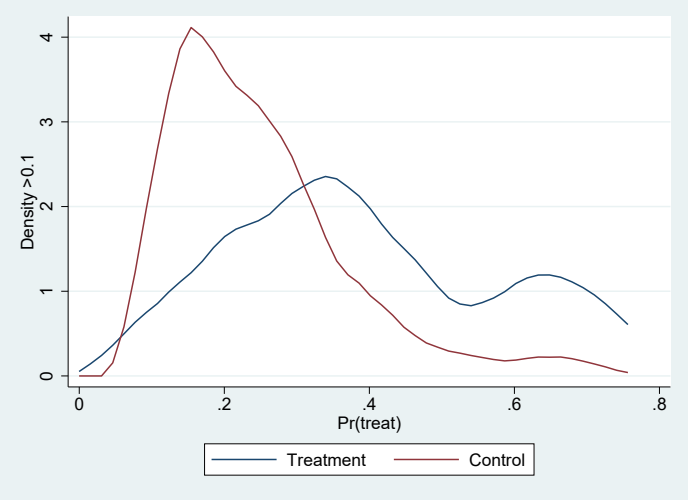

Figure 2: Density

(d)

```
. quietly regress re78 treat age age_2 ed nodeg married black hisp re74 re75  if obs == 0 & pro_
> prop > 0.1, r
. eststo d_reg_1
. quietly regress re78 treat age age_2 ed nodeg married black hisp re74 re75  if obs == 1 & pro_
> prop > 0.1, r
. eststo d_reg_2
. quietly regress re78 treat age age_2 ed nodeg married black hisp re74 re75  if obs == 0 & pro_
> prop > 0.2, r
. eststo d_reg_3
. quietly regress re78 treat age age_2 ed nodeg married black hisp re74 re75  if obs == 1 & pro_
> prop > 0.2, r
. eststo d_reg_4
. esttab d_reg_1 d_reg_2 d_reg_3 d_reg_4, se
```

	(1) re78	(2) re78	(3) re78	(4) re78
treat	1943.8*	1463.1*	2268.8**	1740.2*
	(782.0)	(714.4)	(844.7)	(813.2)
age	32.73	114.1	33.09	148.6
	(360.9)	(342.5)	(419.4)	(495.0)
age_2	0.637	-1.973	0.686	-3.711
	(6.094)	(5.682)	(7.245)	(8.567)
ed	494.8*	324.9*	433.9	310.8
	(218.4)	(132.6)	(239.0)	(195.2)
nodeg	453.1	57.75	-298.1	-1025.9
	(1225.2)	(950.8)	(1627.7)	(1556.9)
married	-78.79	-503.9	-574.6	60.13
	(1054.4)	(852.1)	(1317.5)	(1286.7)
black	-5020.1***	2054.5	0	0
	(686.1)	(1211.8)	(.)	(.)
hisp	2847.8	6206.8	0	0
	(3645.6)	(3471.8)	(.)	(.)
re74	0.0914	-0.0343	0.0729	-0.0590
	(0.112)	(0.0943)	(0.200)	(0.136)
re75	-0.143	0.716***	-0.0109	0.707***
	(0.162)	(0.122)	(0.248)	(0.188)
_cons	2363.1	-2999.7	-1553.9	321.6
	(6626.3)	(6008.6)	(7703.8)	(7691.2)
N	356	502	303	324

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

Columns (1) and (2) are the same regression on experimental data and observational data respectivley, restricted to propenisty score greater than 0.1. Columns (2) and (3) repeat the same thing for propenisty score greater than 0.2.

It seems that we are loosing a big part of the sample and the effects are still similar.

(e)

```
. *teffects psmatch (re78) (treat age age_2 ed nodeg married black hisp re74 re75, logit) if obs
>  == 0 & log_prop > 0.1, vce(robust)
. *eststo e_reg_1
. teffects psmatch (re78) (treat age age_2 ed nodeg married black hisp re74 re75, probit) if obs
>  == 0 & pro_prop > 0.2, vce(robust)
note: black omitted because of collinearity
note: hisp omitted because of collinearity
Treatment-effects estimation                   Number of obs      =        303
Estimator      : propensity-score matching     Matches: requested =          1
Outcome model  : matching                                  min =          1
Treatment model: probit                                    max =          8
```

re78	Coef.	AI Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
treat (1 vs 0)	2239.205	1031.912	2.17	0.030	216.6953	4261.715

```
. eststo e_reg_2
. esttab e_reg_2, se
```

	(1) re78
ATE	
r1vs0.treat	2239.2*
	(1031.9)
N	303

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

This is not feasible for the full sample b/c we have so many ppl close to 0 propensity scroe. For some reason I couldn't get the first regression to work I tried logit and probit, both refused to work. Other than that the results are similar.

Question 3

Import Data

```
. clear all
. use "IPUMS.dta"
.
. *clean data*
. drop if (year == 1960 & hrswork2 <= 1) | (year != 1960 & uhrswork <= 14)
(4,693,543 observations deleted)
. replace uhrswork = 22 if year == 1960 & hrswork2 == 2
(189,462 real changes made)
. replace uhrswork = 32 if year == 1960 & hrswork2 == 3
(122,703 real changes made)
```

```
. replace uhrswork = 37 if year == 1960 & hrswork2 == 4
(205,807 real changes made)
. replace uhrswork = 40 if year == 1960 & hrswork2 == 5
(1,304,713 real changes made)
. replace uhrswork = 44 if year == 1960 & hrswork2 == 6
(538,926 real changes made)
. replace uhrswork = 54 if year == 1960 & hrswork2 == 7
(252,038 real changes made)
. replace uhrswork = 60 if year == 1960 & hrswork2 == 8
(287,878 real changes made)
.
. drop if wkswork2 == 0
(57,551 observations deleted)
. replace wkswork1 = 7 if year == 1960 & wkswork2 == 1
(88,530 real changes made)
. replace wkswork1 = 19 if year == 1960 & wkswork2 == 2
(142,221 real changes made)
. replace wkswork1 = 33 if year == 1960 & wkswork2 == 3
(225,725 real changes made)
. replace wkswork1 = 43.5 if year == 1960 & wkswork2 == 4
variable wkswork1 was byte now float
(259,325 real changes made)
. replace wkswork1 = 48.5 if year == 1960 & wkswork2 == 5
(175,594 real changes made)
. replace wkswork1 = 51.5 if year == 1960 & wkswork2 == 6
(1,952,581 real changes made)
.
. gen log_hr_wage = log(incwage+1/(wkswork1*uhrswork))
.
. gen white = (race == 1)
. replace white = . if (race != 1 & race != 2)
(643,463 real changes made, 643,463 to missing)
. gen female = (sex == 2)
```

(a)

```
. quietly regress log_hr_wage female [pweight = perwt] if year == 2019, r
. eststo a_reg_1
. quietly regress log_hr_wage female age i.educ [pweight = perwt] if year == 2019, r
. eststo a_reg_2
. quietly regress log_hr_wage white [pweight = perwt] if year == 2019, r
. eststo a_reg_3
. quietly regress log_hr_wage white age i.educ [pweight = perwt] if year == 2019, r
. eststo a_reg_4
.
. esttab a_reg_1 a_reg_2, label se
```

	(1) log_hr_wage	(2) log_hr_wage
female	-0.0453***	-0.159***
	(0.00846)	(0.00837)

age		0.000721*
		(0.000312)
n/a or no schooling		0
		(.)
nursery school to ˜4		-0.244*
		(0.109)
grade 5, 6, 7, or 8		-0.524***
		(0.0726)
grade 9		-0.492***
		(0.0801)
grade 10		-0.409***
		(0.0665)
grade 11		-0.222***
		(0.0596)
grade 12		0.503***
		(0.0526)
1 year of college		0.694***
		(0.0531)
2 years of college		1.037***
		(0.0536)
4 years of college		1.448***
		(0.0526)
5+ years of college		1.919***
		(0.0528)
Constant	9.600***	8.718***
	(0.00633)	(0.0538)
Observations	1457373	1457373

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

. esttab a_reg_3 a_reg_4, label se

	(1)	(2)
	log_hr_wage	log_hr_wage
white	-0.0856***	-0.190***
	(0.0123)	(0.0123)
age		0.00155***
		(0.000332)
n/a or no schooling		0
		(.)
nursery school to ˜4		-0.313*
		(0.143)
grade 5, 6, 7, or 8		-0.704***
		(0.0929)
grade 9		-0.465***
		(0.0971)
grade 10		-0.465***
		(0.0809)
grade 11		-0.256***
		(0.0736)
grade 12		0.484***

		(0.0674)
1 year of college		0.645***
		(0.0679)
2 years of college		1.005***
		(0.0683)
4 years of college		1.414***
		(0.0674)
5+ years of college		1.838***
		(0.0677)
Constant	9.664***	8.814***
	(0.0113)	(0.0689)
Observations	1256989	1256989

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

It seems that the female wage gap only grows. And I mustve made a mistake somewhere b/c it seems that blacks are making more than whites?!?

(b) It seems from the regression that they are. I don't know how to interpret for blacks (b/c everything there is backwards), but for female it seems that there might also be selection into eductaion.

(c)

```
. quietly regress log_hr_wage female age i.educ i.occ1990 [pweight = perwt] if year == 2019, r
. eststo c_reg_1
. quietly regress log_hr_wage white age i.educ i.occ1990 [pweight = perwt] if year == 2019, r
. eststo c_reg_2
. esttab c_reg_1 c_reg_2, keep(white female) se
```

	(1) log_hr_wage	(2) log_hr_wage
female	-0.131***	
	(0.00885)	
white		-0.156***
		(0.0120)
N	1457373	1256989

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

This is the regression with occipational controls, it seems there hasn't been much change (although again the coeeficient on white is weird).

(d) Year = 1990

```
. quietly regress log_hr_wage female [pweight = perwt] if year == 1990, r
. eststo d_reg_1
. quietly regress log_hr_wage female age i.educ [pweight = perwt] if year == 1990, r
```

```
. eststo d_reg_2
. quietly regress log_hr_wage female age i.educ i.occ1990 [pweight = perwt] if year == 1990, r
. eststo d_reg_3
. quietly regress log_hr_wage white [pweight = perwt] if year == 1990, r
. eststo d_reg_4
. quietly regress log_hr_wage white age i.educ [pweight = perwt] if year == 1990, r
. eststo d_reg_5
. quietly regress log_hr_wage white age i.educ i.occ1990 [pweight = perwt] if year == 1990, r
. eststo d_reg_6
.
. esttab d_reg_1 d_reg_2 d_reg_3, label se keep(female)
```

	(1) log_hr_wage	(2) log_hr_wage	(3) log_hr_wage
female	-0.103***	-0.123***	-0.248***
	(0.00345)	(0.00343)	(0.00397)
Observations	6098453	6098453	6098453

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

```
. esttab  d_reg_4 d_reg_5 d_reg_6, label se keep(white)
```

	(1) log_hr_wage	(2) log_hr_wage	(3) log_hr_wage
white	-0.335***	-0.441***	-0.256***
	(0.00457)	(0.00453)	(0.00440)
Observations	5679181	5679181	5679181

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

Year = 1960

```
. quietly regress log_hr_wage female [pweight = perwt] if year == 1960, r
. eststo d_reg_7
. quietly regress log_hr_wage female age i.educ [pweight = perwt] if year == 1960, r
. eststo d_reg_8
. quietly regress log_hr_wage female age i.educ i.occ1990 [pweight = perwt] if year == 1960, r
. eststo d_reg_9
. quietly regress log_hr_wage white [pweight = perwt] if year == 1960, r
. eststo d_reg_10
. quietly regress log_hr_wage white age i.educ [pweight = perwt] if year == 1960, r
. eststo d_reg_11
. quietly regress log_hr_wage white age i.educ i.occ1990 [pweight = perwt] if year == 1960, r
. eststo d_reg_12
.
. esttab d_reg_7 d_reg_8 d_reg_9, label se keep(female)
```

	(1) log_hr_wage	(2) log_hr_wage	(3) log_hr_wage
female	0.278***	0.242***	-0.772***
	(0.00605)	(0.00606)	(0.00770)
Observations	2843976	2843976	2843976

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

```
. esttab  d_reg_12 d_reg_11 d_reg_12, label se keep(white)
```

	(1) log_hr_wage	(2) log_hr_wage	(3) log_hr_wage
white	-0.137***	-0.421***	-0.137***
	(0.00794)	(0.00862)	(0.00794)
Observations	2820236	2820236	2820236

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

Column (1) is a regression with no controls, (2) with education controls and (3) with occupation controls