Assignment 2

Sabrina Dorn

Martina Fink

10th March 2010

The data set taken form the National Longitudinal Survey of Youth contains information on ability, education, potential experience, log of hourly wage, mother's and father's education, number of siblings and residence in a broken home for 2,178 individuals. Multiple observations per individual at different points in time give the panel structure of the data set. Overall the data set consists of 17,919 observations.

1. Problem 1

In order to study the relationship between wages and available characteristics in the data set, we run an OLS regression of log hourly wages on education, experience, ability, mother's education, father's education, number of siblings and residence in a broken home. The results are reported in Table 1 column 1.
The overall test rejects that all coefficients except the constant term are zero[1]. Also the joint hypotheses that the coefficients on the four household variables - mother's and father's education, number of siblings and residence in a broken home - can be rejected.[2]

2. Problem 2

2.1. a

The results from the full regression can be found in Table 1 column 2. Education has a highly significant positive effect on log hourly wage of 0.071. Since eduction enters the

[1] $F(7, 17911) = 546.55$ with $p - value < 0.0001$
[2] $F(4, 17911) = 14.03$ with $p - value < 0.0001$

function linearly this effect is constant. By taking $\exp\{0.071\} = 1.07$ one gets the \$/h increase of an one year increase in years of schooling.

2.2. b

The histogram of education shows two pikes at 12 and 15 years. This nonlinearity is incorporated by replacing years of education with two dummy variables indicating a college degree and a graduate degree. Both dummy variables are highly significant. Having a college degree increases log hourly wage by 0.175 compared to no high school graduation. A graduated degree increases log hourly wage by 0.362.

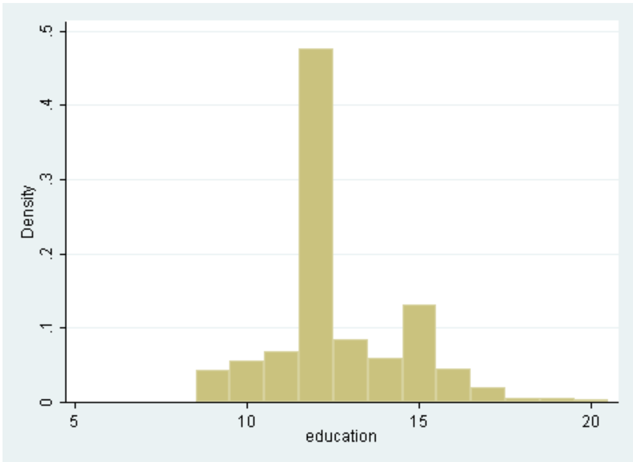

Figure 1: Histogram of education

2.3. c

In model 3 (Table 1 column 3) we incorporate the nonlinearity in the effect of education through a quadratic term. The significant negative coefficient[3] on education squared suggests a decline in the positive effect of education on log hourly wages when education increases.

[3] $p - value < 0.001$ for the t-test with H_0 : coefficient of $education^2 = 0$

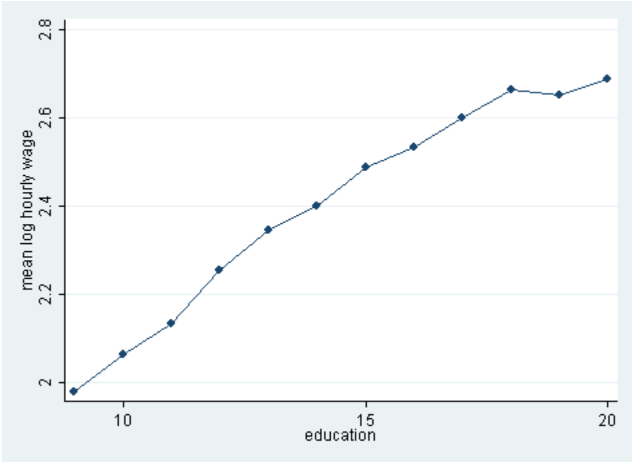

Figure 2: Profile of log wages as a function of education

2.4. d

Model 4 (Table 1 column 4) incorporates the feature that the value of education is enhanced by greater ability. This is done by adding and interaction term of education and ability to the regression. We find a slightly positive but insignificant impact[4] of this interaction. This result suggests that the value of education is the same for individuals with different ability. The marginal effect of education on log hourly wage at mean ability is 0.0702. The corresponding 95% confidence interval is [0.0696, 0.0708].

2.5. e

Finally, the 5th model (Table 1 column 5) combines the nonlinearity in the effect of education and the possible interaction of education and ability. Again we find a declining effect of education as education increases. Now the interaction of education with ability turns out to be significantly positive. Thus the value of education increases with ability.

[4] $p - value = 0.213$ for the t-test with H_0 : coefficient of $education * ability = 0$

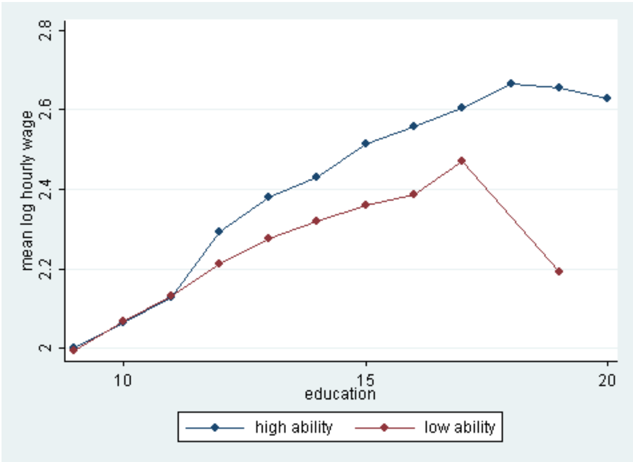

Figure 3: Profile of log wages as a function of education by ability

A. Tables

	Model 1 b/t	Model 2 b/t	Model 3 b/t	Model 4 b/t	Model 5 b/t
education	0.071***		0.156***	0.070***	0.241***
	(31.54)		(8.90)	(28.81)	(10.70)
experience	0.040***	0.038***	0.040***	0.039***	0.040***
	(43.97)	(42.08)	(44.11)	(43.91)	(44.02)
ability	0.077***	0.101***	0.074***	0.047	-0.125***
	(15.68)	(20.75)	(14.96)	(1.88)	(-3.71)
mothers education	0.000	0.001	0.000	0.000	0.000
	(0.04)	(0.48)	(0.18)	(0.03)	(0.27)
fathers education	0.005***	0.007***	0.005***	0.005***	0.005***
	(3.97)	(5.19)	(3.88)	(4.00)	(3.93)
broken home	-0.053***	-0.070***	-0.050***	-0.053***	-0.048***
	(-5.29)	(-6.91)	(-4.95)	(-5.32)	(-4.77)
number of siblings	0.005**	0.004*	0.005**	0.005**	0.005*
	(2.72)	(2.05)	(2.79)	(2.67)	(2.57)
college		0.175***			
		(20.02)			
graduate		0.362***			
		(17.37)			
education squared			-0.003***		-0.007***
			(-4.88)		(-7.63)
education*ability				0.003	0.016***
				(1.24)	(5.99)
cons	0.990***	1.811***	0.428***	1.002***	-0.105
	(29.20)	(87.52)	(3.56)	(28.39)	(-0.70)
adj. R^2	0.176	0.157	0.177	0.176	0.178
N	17919	17919	17919	17919	17919

Table 1: Results

Note: * $p<0.05$, ** $p<0.01$, *** $p<0.001$

B. STATA code

```
clear

cd C:\Users\fink\Desktop\phd\econometrics\assignment2\

cap log c
log using data, replace

* IMPORT TO STATA

infile abil moth fath brok sibl using time_invar.dat
gen id = _n
la var abil "ability"
la var moth "mothers education"
la var fath "fathers education"
la var brok "broken home"
la var sibl "number of siblings"
la var id "person id"
save time_invar.dta, replace

infile id educ wage exp time using time_var.dat, clear
la var id "person id"
la var edu "education"
la var wage "log hourly wage"
la var exp "experience"
la var time "time trend"
save time_var.dta, replace

mmerge id using time_invar.dta, type(n:1)
drop _merge

save data.dta, replace


* SUMMARY
su *
```

```
* PROBLEM 1

loc X1 "educ exp abil"
loc X2 "moth fath brok sibl"

* 1.A.
reg wage `X1' `X2'
estimates store m1, title(Model 1)

* 1.B.
* see reg output from 1.A.

* 1.C.
test (`X2')


* PROBLEM 2

loc X "`X1' `X2'"

* 2.A.
reg wage `X'


* 2.B.
histogram educ, discrete
gr export histogram.png, replace
gen byte hs = educ <= 12
la var hs "high school"
gen byte col = educ > 12 & educ <= 16
la var col "college"
gen byte grad = educ > 16
la var grad "graduate"

reg wage col grad exp abil moth fath brok sibl
estimates store m2, title(Model 2)
```

```
* 2.C.
gen educ2 = educ^2
la var educ2 "education squared"
reg wage educ educ2 exp abil moth fath brok sibl
estimates store m3, title(Model 3)
predict temp, xb
bys educ: egen temp2 = mean(temp)
la var temp2 "mean log hourly wage"
bys educ: gen temp3 = _n == 1
twoway (connected temp2 educ if temp3 == 1)
gr export profile1.png, replace
drop temp temp2 temp3

* 2.D.
gen educXabil = educ * abil
la var educXabil "education*ability"
reg wage educ educXabil exp abil moth fath brok sibl
estimates store m4, title(Model 4)
test (educXabil)
mat b = e(b)'
mat w = (1\0.052374\0\0\0\0\0\0\0)
mat c = w'*b
mat vc = e(V)
mat var_margeff = w'*vc*w
sca lower = c[1,1]-tden(_N,0.975)*sqrt(var_margeff[1,1])
sca upper = c[1,1]+tden(_N,0.975)*sqrt(var_margeff[1,1])
di "marg. eff.t for add. year of education with mean ability: " c[1,1]
di "corresponding conf. interval: [" lower "," upper "]"


* 2.E.
reg wage educ educ2 educXabil exp abil moth fath brok sibl
estimates store m5, title(Model 5)
su abil, meanonly
gen byte highabil = abil>r(mean)
predict temp, xb
```

```
bys educ highabil: egen temp2 = mean(temp)
la var temp2 "mean log hourly wage"
bys educ highabil: gen temp3 = _n == 1
twoway (connected temp2 educ if temp3 == 1 & highabil == 1)
   (connected temp2 educ if temp3==1 & highabil == 0),
   legend(lab(1 "high ability") lab(2 "low ability"))
gr export profile2.png, replace
drop temp temp2 temp3

estout *, style(tex) cells(b(star fmt(3)) t(par fmt(2))) legend label stats(r2_a N)


log c

exit
```