Assignment 1

Walter Verwer & Bas Machielsen

1/3/2021

Question 1

1. The sample selection model. A researcher aims to gain insight in the potential earnings of the non-employed. (In the data, the non-employed can be identified by a missing value for the earnings variable). She realizes that the sample of observed wages may be subject to sample selection.

(a) Run an OLS regression for log-earnings on schooling, age, and age squared. Present the results and comment on the estimates.

```
model <- lm(data = data, formula = logWage ~ schooling + age + age2)
stargazer(model, type = "html")
```

Dependent variable:

logWage

schooling

0.216***

(0.032)

age

-0.342

(0.521)

age2

-0.011

(0.008)

Constant

26.409***

(8.057)

Observations

416

R2

0.815

Adjusted R2

0.813

Residual Std. Error

1.499 (df = 412)

F Statistic

604.261*** (df = 3; 412)

Note:

p<0.1; ***p<0.05;*** p<0.01

The results show that 1 additional year of schooling has an effect of 0.216 on log(Wage), which means that it has an effect of 1.2411024 on schooling. Hence, each year of schooling supposedly increases earnings with 1.2411024.

Hoi bas dit is een test. Hieronder doe ik code, wat ik wil laten zien:

```
x <- 3*3 + 1
```

(b) The researcher has collected data on two potential instrumental vari- ables subsidy and distance for years of schooling.

- distance measures the distance between the school location and the residence of the individual while at school-going age.
- subsidy is an indicator depending on regional subsidies of families for covering school expenses.

The researcher has the option to use only distance as an instrumen- tal variable, or to use only the instrumental variable subsidy, or to use both distance and subsidy as instrumental variables. Perform instrumental variables estimation for these three options. Which option do you prefer? Include in your answer the necessary analyses and numbers on which you base your choice.

```
firstoption <- ivreg(data = data, formula =
                          logWage ~ age + age2 + schooling | distance + age + age2)

secondoption <- ivreg(data = data, formula =
                           logWage ~ age + age2 + schooling | subsidy + age + age2)

thirdoption <- ivreg(data = data, formula =
                           logWage ~ age + age2 + schooling | subsidy + distance + age + age2)
```

```
stargazer(firstoption, secondoption, thirdoption)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sun, Jan 03, 2021 - 15:40:10

Table 1:

	Dependent variable:		
	logWage		
	(1)	(2)	(3)
age	−0.192	−0.233	−0.229
	(0.587)	(0.546)	(0.547)
age2	−0.014	−0.013	−0.013
	(0.010)	(0.009)	(0.009)
schooling	0.470	0.401***	0.408***
	(0.299)	(0.106)	(0.102)
Constant	22.681**	23.694***	23.589***
	(9.704)	(8.517)	(8.530)
Observations	416	416	416
R^2	0.786	0.799	0.798
Adjusted R^2	0.784	0.798	0.797
Residual Std. Error (df = 412)	1.613	1.560	1.565
Note:			*p<0.1; **p<0.05; ***p<0.01