ECON 672 Winter 2022 Problem Set #1

Jon Holder

January 11, 2022

Question 1

a. Persico, Postlewaite, and Silverman (2004) are examining the mechanism that influences the relationship between height and earnings gap. They note that previous studies have evaluated the influence of current height. They take advantage of the fact that height changes throughout youth before becoming relatively constant in adulthood. Persico, Postlewaite, and Silverman find that height at 16 is a predictor of future earnings controlling for vocation, family attributes and adult height.

b. Two data sets are evaluated: 1) Britain's National Child Development Survey (NCDS) and the US's National Longitudinal Survey of Youth (NLSY) 1979 cohort.

c. Table 1 shows the raw heights, change in heights and their distributions for white men in the two data sets. Persico, Postlewaite, and Silverman include this data because it is the key parameter of interest. The characteristics and distributions of these heights provide important context for the conclusion of the authors.

d. Table 2 shows various statistics for sample. It includes traditional determinants of wages include individual and family educational attainment, martial status, and number of siblings. Finally, Table 2 contains the ln(wages) which is the dependent variable.

e. Table 3 shows various OLS estimates for effect on ln(wages) using the controls from Table 3. The estimates show a positive return to ln(wages) for full-time white male workers based on reported height during youth.

f. Table 4: **TODO**

g. Summary: **TODO**

Question 2

b. (i) NLSY contained 3 subsamples: a cross-sectional sample that represented noninstitutionalized civilian population, a sample that oversampled civilian Latino, black and other economically disadvantage civilians, and a sample representing the population that was in the military. After the 1984, most members of the military where ineligible for interview so 201 members of that cohort were randomly selected to remain.

After the 1990 interview, the economically disadvantaged white population was removed from the sample as noted in the analysis by Persico, Postlewaite, and Silverman (2004).

(ii) The survey collected information about the respondents family background and family formation. They collected information about the respondents education background including high school, college, and/or vocational training as well as a standardized test of general ability, ASVAB. Finally, the survey collected information about labor participation, government program participation, and income and assets.

f. Height data is available in 1981, 1982,1985, 2006, 2008, 2010 2012, 2014. 2016, 2018.

Question 4

a. There are 12,686 observations in the dta file.

b. There in are 6,403 (50.47%) males and 6,283 (49.53%) females in the 1979 sample.

Question 5

Restrict sample to non-poor white men.

Question 6

a. Generate **height81**

(i) HEALTH_HEIGHT_1981 is encoded as height in feet and inches displayed as a 3 digit integer value.

(ii) -5 response means that the person was not interviewed.

(iii) Responses below 0 are error codes. Additionally height cannot be negative.

(iv) A response of "510" means that respondent is 5 feet and 10 inches tall.

(v) *Generate Stata variable*

(vi) The mean height for white men in 1981 is 70.34 inches or about 5ft 10.3in.

b. Generate **height85**

(i) *Generate Stata variable*

(ii) The mean height of a white man in the 1985 sample is 70.64 inches or 5ft 10.65inches.

(iii) Error coded observations were dropped. Values less than 0 are error coded.

(iv) The mean change in height between 1985 and 1981 is .2889 inches.

c. Generate **age96**

(i) FAM_1B_1979 is the respondent's age in 1979.

(ii) *Generate Stata variable*

d. Generate **income**

(i) Income in truncated for the top 2%. Negative values are error codes but 0 is valid value.

(ii) *Generate Stata Variable*

(iii) There are 2,094 valid observations for income.

(iv) The mean of the valid income is $34,912.86.

e. Generate **hours**

(i) Negative values are error codes but 0 is valid value.

(ii) *Generate Stata Variable*

(iii) There are 2,139 valid observations for hours in the dataset.

(iv) The mean hours worked is 2,222.4.

(v) Generate *fulltime* variable

(vi) 90.04% of the remaining respondents worked fulltime.

f. Generate **lnWage**

(i) *Generate Stata Variable*

(ii) There are 1,973 observations that have non-missing lnWage values.

(iii) The remaining sample size with non-missing values is 1,979.

g. Generate **educ**. This is the highest level of education reported from the 1979,1981, 1985, and 1996 surveys.

(a) Missing values are negative.

(b) *Generate Stata Variable*

(c) The mean number years completed is 12.2 yrs. This corresponds to a little more than a high school

h. Generate **everMarried**

(a) Values less than 0 are error codes. Values 1 or higher indicate the respondent is currently or was previously married and is now either separated, divorced or widowed.

(b) *Generate Stata Variable*

(c) About 78.48% of the sample have been married.

i. Generate **momSchool**

(i) Error codes are negative. A value of 95 means the respondent has an ungraded highest level of education. However, no respondents have this value.

(ii) *Generate Stata Variable*

(iii) The mean level of the mother's education for the sample is 11.9 years.

j. Generate **momSkilled**

(i) This contains classification codes. Codes 1-245 are considered professional/managerial. Negative values are error codes.

(ii) *Generate Stata Variable*

(iii) The share of mothers in skilled occupations in 1979 is 19.7%.

k. Generate **dadSchool**. Generate **dadSkilled**

(i) The mean level of school for dads is 12.2 years.

(ii) The share of dads in skilled occupations in 1979 is 32.2%

l. Generate **siblings**

m. Generate **finalSample**

(i) *Generate Stata Variable*

(ii) The final sample size is 910 responses.

Question 7

	Mean	Median	Standard Deviation	25^{th} Percentile	75^{th} Percentile	Observations
A. United States NLSY Entire Sample						
Height 1981	70.34	71	2.82	69	72	2900
Height 1985	70.64	71	2.78	69	72	2343
Δ1981-1985	0.29	0	1.44	0	1	2281
B. Final Estimation Sample						
Height 1981	70.36	71	2.90	69	72	910
Height 1985	70.74	71	2.75	69	73	910
Δ1981-1985	0.37	0	1.48	0	1	910

Table 1: *Replication of Persico, Postlewaite, and Silverman (2004) Table 1*

Question 8 TODO

	Adult Height Median or Below (1)	Adult Height Above Median (2)	Difference (3)
Adult Characteristics			
1981 Height	68.61	72.56	3.95*
	(2.31)	(1.92)	(0.14)
1985 Height	68.83	73.12	4.285*
	(2.03)	(1.28)	(0.11)
Age	34.45	34.48	0.02
	(2.26)	(2.32)	(0.15)
ln(wage/hour)	2.63	2.72	0.09*
	(0.65)	(0.64)	(0.04)
Ever Married(%)	85.57	89.11	0.03
	(35.17)	(31.19)	(0.02)
Family Background			
Mother's years of schooling	12.26	12.44	0.19
	(2.43)	(2.13)	(0.15)
Mother skilled\professional (%)	18.77	21.29	0.02
	(39.09)	(40.98)	(0.03)
Father's years of schooling	12.51	12.87	0.36
	(3.20)	(3.09)	(0.21)
Father skilled\professional (%)	32.41	35.15	0.03
	(46.85)	(47.80)	(0.03)
Observations	506	404	910

Table 2: *Replication of Persico, Postlewaite, and Silverman (2004) Table 2.* $*p < 0.05$

References

Persico, Nicola, Andrew Postlewaite, and Dan Silverman. 2004. "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height." *Journal of Political Economy* 112 (5): 1019–1053. ISSN: 00223808, 1537534X. http://www.jstor.org/stable/10.1086/422566.