WRITTEN PAPER II (ECON 4135)

October 4, 2006

Deadline for handing in: October 18, 2006

The relationship between schooling and earnings is one of the most frequently studied one in empirical economics. A large part of these studies are versions of the earnings function setup proposed by Mincer (1974). A key parameter of the Mincer earnings relation is the coefficient associated with years of schooling. This coefficient is meant to capture the marginal economic return to an additional year of schooling.

The main purpose of this exercise is to examine the relationship between log earnings, $\ln Y$, where Y is total annual taxable labor income, and years of schooling, S. Two STATA data sets accompany this problem set: earningsdata_males.dta and earningsdata_females.dta. The data sets contain individual level information about log earnings, years of schooling, labor market experience, type of education, sector of occupation and work place region ("fylke"). The observation units in earningsdata_males.dta are 5859 native-born males who lived in Norway in 1970, was born between 1952 and 1970, and who still lived in Norway in 1997. The data set earningsdata_females.dta contains the same variables for 3247 females.

The samples (males and females) are randomly drawn form the Norwegian system of register data in 1997 and is restricted to full-time wage-earners, defined as individuals working 30 hours or more per week. The measure of years of schooling is defined as the standard number of years necessary to complete this level. Labor market experience is represented as age minus years of schooling minus 7 years, i.e. *potential* (not *actual)* experience. The earnings measure reflects annual earnings, observations whose employment relationships started or terminated within the actual year were excluded. Holders of multiple jobs and individuals that received labor market compensation or have participated in active labor market programs have been excluded. A full list of variables is given in Table 1.

1 . Use the data set for males and estimate the regression equation

$$\ln Y_i = \alpha + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + u_i. \quad (1)$$

Specify your assumptions about the error term u_i and discuss briefly the implications for the OLS estimator of each of these assumptions. What is the conditional expectation $E(\ln Y_i|S_{i,}E_i)$? Give a 95% confidence interval for β_1. What is your estimated expected marginal return to schooling, $\partial E(\ln Y_i|S_{i,}E_i)\,/\partial S_i$? What is the estimated percentage increase in income of one additional year of schooling?

2. You have probably found that the estimated coefficient β_3 is negative. What is the interpretation of this? Use the mean value of E_i in the sample, $\overline{E} = \frac{1}{n}\sum_{i=1}^{n} E_i$, and find an expression for $\partial E(\ln Y_i|S_{i,}E_i)\,/\partial E_i$ when $E_i = \overline{E}$. How would you test whether $\partial E(\ln Y_i|S_{i,}E_i)\,/\partial E_i = 0$ (for any E_i)? Carry out the test. Test whether there is a linear relation between $\ln Y_i$ and E_i.

3. Now consider the model

$$\ln Y_i = \alpha + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \sum_k \gamma_k X_{ik} + u_i, \qquad (2)$$

where the X_{ik} are the remaining variables in the data set, i.e. not used in equation (1), and the γ_k are the corresponding regression coefficients. All the variables X_{ik} are binary variables, which each take the value zero or one; i.e. they are so-called indicator variables. There are three types of variables: indicators of sector of occupation, indicators of workplace region and indicators of type of education. Is there a multicollinearity problem here in the sense that some of the X_{ik}-variables always sum to one? If your answer is yes: how could you remedy the problem? If your answer is no: what is the basis category (excluded category) for each of these three types of variables?

4. How does your estimate of β_1 change compared to equation (1)? Interpret the results. Do you find evidence of omitted variable bias when you exclude the X_{ik}-variables from the regression?

5. What model specification do you prefer: equation (1) or (2)? State the reasons for your choice. Use your preferred model specification, but now with the data set for females. Compare the results for males and females. Is the parameter β_1 significantly different for males and females? Set up a test statistic and conduct a test: (Hint: Test wether $\beta_1^{male} - \beta_1^{female} = 0$ versus $\beta_1^{male} - \beta_1^{female} \neq 0$, with the obvious meaning of β_1^{male} and β_1^{female}).

6. Can you use STATA or any other program to make graphs of $\partial E(\ln Y_i|S_{i,}E_i)\,/\partial E_i$ as a function of E_i – for both males and females in the same figure. Comment on the differeces between males and females.

7. What is *your* optimal choice of (level of) schooling?

REFERENCES

Mincer, J. (1974). *Schooling, experience and earnings.* New York: National Bureau of Economic Research.

Table 1: **Variable list**

Variable name	Description
lnY	Log of income
S	Years of schooling exceeding 7 years
E	Years of experience
E^2	Years of experience squared
public	Sector of occupation= public services (1 if true, 0 else)
servi	Sector of occupation= private services
unsp	Type of education=unspecified (1 if true, 0 else)
gen	Type of education=general
hum	Type of education=humanities
teach	Type of education=teaching
admin	Type of education=business/adminstrative
transp	Type of education=transport
health	Type of education=health
farm	Type of education=farming/fisheries
serv	Type of education=services/military
ostf	Region=Østfold (1 if true, 0 else)
akershus	Region=Akershus
hedemark	⋮
oppland	⋮
buskerud	⋮
vestfold	⋮
telemark	⋮
a-agder	⋮
v-agder	⋮
rogaland	⋮
hordaland	⋮
sognfj	⋮
moreroms	⋮
s-tr	⋮
n-tr	⋮
nordland	⋮
troms	⋮
finmark	Region=Finmark