Advanced Development Economics
University of Göteborg
Fall 2009
Annika Lindskog & Måns Söderbom

Computer Exercise I

Earnings and Education in Pakistan

Instructions: You may work on this exercise in a group consisting of 2-3 students. Each group needs to submit their solutions in the form of a short **report** to:

Annika.Lindskog@economics.gu.se

or

Mans.Soderbom@economics.gu.se.

The first page of the report should be a cover page with the names of the students clearly indicated. Your report should contain brief discussion of the results, plus show the results themselves (however there is no need to include Stata syntax in the report). There is no formal deadline, but please note that you will not be able to claim the credits for this course until you complete the two computer exercises.

This exercise is based on data from the third round of the Pakistan Integrated Household Survey (PIHS) conducted in 1998-1999.[1] The PIHS provides a nationally representative sample made up of around 16,000 households, which represent roughly 115,000 observations. The household questionnaire is composed of a number of detailed modules on such characteristics as income, education, health, maternity and family planning, consumption and expenses, housing conditions and available services. The sample used throughout this study consists of individuals aged between 16 and 70 and not currently in school (hence the total number of observations is less than 115,000).

The following variables are included in the dataset **pakistan_households98.dta**:

```
Contains data from j:\CE1_ADE09\pakistan_households98.dta
  obs:        47,804
 vars:            11                          21 Sep 2009 14:29
 size:     1,864,356 (99.1% of memory free)
------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
age             byte   %8.0g                  age
educ            float  %29.0g      eduyears   years of education
kidsund12       byte   %9.0g                  Number of kids under 12 in
                                                household
eldove65        byte   %9.0g                  Number of elderly over 65 in
                                                household
readwrit        float  %9.0g                  can read & write
canmath         float  %9.0g                  can solve simple maths problem
married         float  %9.0g                  dummy = 1 if married
male            float  %9.0g                  dummy = 1 if male, = 0 if female
cat             float  %9.0g                  occupation category
lw              float  %9.0g                  log annual earnings in Pakistan
                                                Rupees
wemp            float  %9.0g                  dummy = 1 if wage employee
------------------------------------------------------------------------------
```

All individuals in the labor market are classified into either one of 5 occupational categories: self-employed (cat=1), working in the agricultural sector (cat=2), wage-employee (cat=3), unemployed (cat=4) and out of the labor force (cat=5). Unemployed individuals are those who seek employment and are available for it, while out of labor force

[1] For an analysis of these data, see Kingdon, Geeta and Måns Söderbom, "Education, Skills, and Labor Market Outcomes: Evidence from Pakistan," Education Working Paper Series, no. 11, May 2008. Washington D.C: The World Bank. This can be downloaded at http://www.soderbom.net/ADElab1.pdf. This is not required reading, however.

individuals are those who do not seek employment, such as housewives and the retired.

Two variables measuring cognitive skills are included. Both are based on a self-reported measure of whether the respondent can read and write (literate) and do simple sums (numerate).

Questions

a) What is the average years of education in this sample? How does the average years of education differ between men and women?

b) How do reading and writing skills differ between men and women?

c) Redo the calculation in (a) and (b) for the relatively young; say, those aged between 16 and 25. Compare the results to those reported in (a) and (b). Discuss.

d) Estimate the proportions of individuals belonging to each of the five occupation categories, for men and women separately. Comment on the results. What is the labour force participation rate (LFP), for men and women, respectively?

e) Does acquiring education make it more likely that you get a wage job? Answer this question for men and women, separately.

f) What is the correlation between education and log earnings? What is this telling you regarding the role of education in improving standards of living?

g) Consider the following simple earnings function:

$$\ln w_i = \beta_0 + \beta_1 \cdot educ_i + \beta_2 \cdot age_i + \beta_3 \cdot age_i^2 + \beta_4 \cdot MALE_i + \varepsilon_i$$

where *w* denotes earnings; educ, age, age^2,MALE are explanatory variables; ε is an error term, and *i* is an individual index. Explain how to interpret the β-parameters. [Do not proceed beyond this point until you are clear on they should be interpreted!]

h) Estimate the earnings function above using OLS. Discuss the results. What do the results imply as to the difference between men and women in expected rupees earned, conditional on education and age?

i) Modify the econometric model above so as to enable you to investigate if the returns to education differ between men and women. Explain what you are doing, and why. Obtain estimation results and discuss these.

j) Does being married tend to increase earnings? [Hint: This is a trick question]

k) Modify the econometric model above so as to enable you to investigate if reading, writing and maths skills tend to raise earnings. Obtain OLS results and discuss the effects of these types of skills. Keep an eye on gender differences. Should education be included in the set of explanatory variables?

l) It is often argued that education is correlated with the disturbance term in earnings functions. Is this likely to be a problem in the regressions estimated above? If so, what can be done about it?