B4 - Applied Statistical Methods
Fall 2017
Home Work Assignment: Wage equations

Johan Lyhagen/ Lars Forsberg

1 Handing in

Deadline is in the schedule. Handing in is done in Studentportalen.

1. The name of the attached .pdf file should be:

   "B4-HWA-Wage-Familyname-Familyname-Familyname.pdf"

   where

2. Family names must be in alphabethical order.

3. you replace the text 'Familyname' with the family name of the first group member second etc, as ordered in alphabetical order by family name.

Data

Data comes from the project "Hushållens ekonomiska levnadsvillkor, HUS" (Households economic living conditions). HUS is a database about households use of time, money and public services. HUS also contains background variables, e.g. education, working experience and salaries. The survey is representative for native Swedish speaking families living in Sweden when the survey was conducted. For further information regarding the survey see Klevmarken & Olovsson (1993): "Household Market and Nonmarket Activities: Procedures and Codes 1984-1991".

Data for the HWA comes from HUS 1984 and are available in Datafile "b4 - hwa - wage - data.csv" on Studentportalen. The following variables are in the data file: *birth_year, gender* (1 for male, 2 for female), *x102-x104, x108, x109, x113, x114, x118, x119, x130, x140, x210, x258, x323*. The pdf file "Description of variables.pdf" gives a description of the variables.

Tasks

1. Perform a preliminary analysis of the data. For example, study the mean, variance, min, and max of the variables you intend to use in the tasks below. This preliminary analysis also allows detection of possible wrong values. Generate the variables needed for the tasks below (e.g. log(wage) etc).

2. Specify and estimate a wage model. A very basic wage model is of the form (Mincer's earnings equation)
$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u,$$
where *educ* denotes the education and *exper* the experience measured in working years of a respondent. Do the residuals in the wage model you estimated seem to be heteroscedastic? If yes, employ an approach to deal with that problem. Interpret your results. (Assume that all the explanatory variables of your model are exogenous.)

3. Is there a difference in wages between men and women? If so, is that due to discrimination or does it have other causes?

   (a) First, test if there is a difference in wage. Perform a test for this without formulating a regression model. (Test of difference in means.)

   (b) Formulate a regression model that will allow you to test if there is a difference in wage and perform this test. This model should give the same result as the test in previous subtask. Present the results in an intuitive and easy to understand table.

4. Now, some might say that difference in wage between genders is not due to discrimination, but due to other factors. (Thus the difference that you proved in previous task is not at all due to discrimination.)

   (a) Formulate and estimate a model that takes difference in *education* into account, that is, a model where you control for the difference in education. Given that you have controlled for these differences - is there still a difference? Perform a formal test of the relevant parameter(s). Interpret the results.

   (b) Another argument is that difference in wages is due to, not just differences in education, but also difference in *experience*. Formulate and estimate a model that takes both education and experience (no need to include squared experience here) into account, that is, a model where you control for the difference in education and experience. Given that you have controlled for these differences - is there still a difference? Perform formal test of the relevant parameter(s).

   (c) Now, assuming that Education and experience are the only variables that matters (a huge simplification, but for the sake of argument), this means that any difference that is still there after controlling for difference in education and experience is due to discrimination. You should have concluded that there is gender discrimination - what is the estimated proportion of wage difference due to gender discrimination? (You need to quantify the effect of difference in education and difference in experience and whatever is left over, is the effect of discrimination.) That is, difference in education and difference in experience can explain some, but not all the difference in wage. Use the same model

you estimated in task 2. (No binary variables as explanatory, hint, recall lecture slides on this.) Assume again that all explanatory variables of your model are exogenous, i.e. no need for instruments in the estimation.

5. However, sometimes it is argued that the variable education is correlated with the error term in a wage model, due to the fact that *ability* is not in the model, and that would be correlated with education. If this is the case, OLS is biased and inconsistent and you need to estimate the model using a instrumental variable method. And to do that, you need instruments.

    (a) Describe your (potential) instrumental variables (for education) and discuss your choice of those instruments for this situation.

    (b) Perfom test(s) to make sure that the instruments you plan to use are relevant.

6. Now, re-do Task 4 using 2SLS;

    (a) Re-do Task 4a, that is, estimate the model that takes difference in education into account, but estimate it using 2SLS.

    (b) Re-do Task 4b, that is, estimate the model that takes difference in education and difference in experience into account, but estimate it using 2SLS.

    (c) When estimating the model correctly - what is the estimated proportion of wage difference due to gender discrimination? (You need to quantify the effect of difference in education and difference in experience and whatever is left over, is the effect of discrimination.) That is, redo task 4c but assume that education is endogenous, that is, use 2SLS.

7. Compile the results from task 4 and 5 in an intuitive and easy to understand result table, and comment on the differences between OLS and 2SLS. Is there a difference? If so, how much? Comment on the results.

8. Now, formulate and estimate (again assuming endogeneity) a model such that it is possible to test the hypothesis that one more year of education is more worth to a man than to a woman. Also, given this model, is should be possible to test if one more year of experience is more worth to a man than a woman. Perform relevant test(s), and interpret the relevant parameter(s).