
Problem Set 1: Estimation of an earnings function

The German Socio-economic Panel (G-SOEP) is a longitudinal survey of private households and persons in the Federal Republic of Germany. The central aim of this panel study is to collect representative micro-data on persons, households and families in order to measure stability and change in living conditions by following principally a micro-economic approach enriched with sociology and political science variables. (More information on the data set can be found here: www.diw.de/soep)

In the following, we use data from the 2004 survey which can be found in the file „verdienst04.raw“ on the L-drive. It contains variables on earnings, education, and work experience for people living in West Germany with completed education under the age of 65 who work full time. Read in the data file and answer the following questions.

The data set contains the following variables in the given order:

Variable name	Variable description
sex	Dummy for sex: 1=woman, 0=man
age	Age in years
educ	Years of education
exper	Work experience in years
noschool	Dummy for no school degree
medschool	Dummy for low or medium school degree
highschool	Dummy for high school degree
notrain	Dummy for no training degree
appren	Dummy for apprenticeship
uni	Dummy for university degree
earn	Gross monthly earnings

1. Estimation of a simple Mincerian earnings function (multivariate OLS)

- a) Regress the earnings on the years of education and on work experience and interpret your results. What are the marginal effects?
Hint: The TSP-command for ordinary least squares is: `olsq y c x1 x2 ...;`
- b) Are the variables individually and/ or jointly significant? (t-test and F-test)

2. Extension of the Mincerian earnings function (Dummy variables)

- a) Do women earn less than men? By how much? In order to estimate this, add a sex dummy to the regression and interpret the results. What exactly does the coefficient of the dummy variable tell us?
- b) Does the effect of an extra year of schooling differ between women and men? In order to check that, include an appropriate interaction term in the specification. What is the marginal effect of education on earnings?

Hint: In order to generate a new variable in TSP use: `genr z = expression ;`

- c) A criticism of the Mincerian earnings function states that earnings do not depend that much on years of education but instead on the highest obtained degree. Which assumption is implicitly included in a linear specification of years of schooling? How can this assumption be loosened? And how can it be tested?
Include appropriate dummy variables in the specification and interpret the results. Moreover, test the hypotheses that these newly added variables are relevant.

3. Specification

- a) Estimate the earnings function with a second order polynomial in the work experience. Based on hypothesis tests, explain which specification should be preferred here.

Hint: You have to generate the second order polynomial, for example like this:
`genr exper2 = exper**2 ;`

- b) Based on the quadratic specification in work experience, what is the effect of work experience on earnings? By how much do earnings increase for another year of work experience
 - i) after the first year on the job?
 - ii) after ten years of work experience?

4. Supplementary exercise (semilog model)

Usually the natural logarithm of earnings is used. Therefore, now use logarithmized earnings and interpret the coefficients correctly.