UNIVERSITÄT SIEGEN

School of Economic Disciplines
Chair for European Economic Policy

Vertr.prof. Dr. Lars-H. R. Siemers   Summer Term 2016

Lecture "Introduction to Stata"
Problem Set 2

What determines the salaries of CEOs?

Open the data set CEO_salary.dta. It provides observations on 12 variables of firms:

1. salary: 1990 annual salary (in 1,000 $)
2. pcsalary: % change salary 1989-1990
3. sales: 1990 firm sales (in million $)
4. roe: average return on equity, 1988-1990 (net income in relation to common equity in %)
5. pcroe: % change roe 1988-1990
6. ros: return on firm's stock, 1988-1990
7. indus: 1 for industrial firm, 0 otherwise (dummy variable)
8. finance: 1 for financial firm, 0 otherwise (dummy variable)
9. consprod: 1 for consumer product firm, 0 otherwise (dummy variable)
10. utility: 1 for transport or utilities, 0 otherwise (dummy variable)
11. lsalary: natural logarithm of salary
12. lsales: natural logarithm of sales

We are interested in the research question "What determines the salaries of CEOs?" The data contain information on salaries in 1990 in thousand US-$. Solve the following problems. Use again a do and log file.

a) Scatterplot salary and roe to get a first impression how the salaries may interact with the rate of return to equity. Which relationship do you conclude from the scatterplot? Do you worry about outliers?

b) Now add the bivariate regression line to the scatterplot. Then, exclude the two observations with salaries above 10 million US-$ and reproduce the graphic. Do the two observations affect the result markedly?

c) Regress salary on roe by OLS. Interpret R^2 and check whether the coefficient of roe is statistically significant (different from zero). Then exclude the two observations with salary $> 10,000$ and repeat the regression. What is the effect? What do we conclude?

d) Maybe there is a difference between sectors. Check this possibility by regressing our model for industrial firms, financial firms, consumer product firms, and firms in the sector transport or utilities separately. Have a look at the respective p-values of the slope parameter. What do you conclude?

e) Is it the change of roe in the years 1989 to 1990 that determines the change of the salary in the same time period? Are there differences between the sectors?

f) Check whether the sales influence the salaries. Also use the specification where the log variables of the salary and the sales are used, instead of their levels.

g) Depict the scatterplot inclusive the regression line for the level and for the log version. Compare the figures and interpret.

h) Instead of simple OLS regressions now use a multiple OLS regression: Regress lsalary on lsales, roe and ros. Interpret the coefficients. What do think about this specification?

i) Now add the industry dummies to the specification (industry fixed effects). Why is one dummy dropped? What is the correct interpretation?

j) Reproduce the F-test for joint significance and the t-test for lsales for the former regression via the command `test`.

k) Now exclude ros (why should we do this?) and repeat the regression for this specification. Compare R-squared and adjusted R-squared of the specifications before and here. Have a look on the F statistics and its p-value, too.

l) Now generate interaction terms between the industry dummies and lsales. Include them to the regression model and re-estimate. What is the interpretation of their coefficients? Is there an effect on the interpretation of the other coefficients?