R TUTORIAL 6

Siria Angino
Federica Romei

1. The dataset *ceosal1.dta* contains 209 observations for the year 1990 on the the following data:

   - *salary*: 1990 CEO salary, thousands $;
   - *sales*: 1990 firm sales, millions $.
   - *roe*, 1990 return on equity.
   - four dummy variables: *indus, finance, consprod, utility,* equal to 1 if the firm is an industrial, financial firm, consumer product or transportation/ utilities firm respectively, 0 otherwise.

   (a) What is the percentage difference in salary for a CEO working in the financial sector and another working in the industrial sector, holding all other things constant?

   Solution:

   Since we are asked for a percentage difference, we apply the logarithmic transformation to the variable "salary". Then we omit "indus" or "finance" for convenience. The coefficient of the non-omitted variable will represent the difference we are looking for (otherwise, we should take the difference between the two coefficients).

```
logsalary=log(salary)
summary_rob(lm(logsalary~sales+roe+finance+consprod+utility))
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.677e+00  1.190e-01  56.118  < 2e-16
sales        1.369e-05  7.767e-06   1.763 0.079411
roe          8.413e-03  4.654e-03   1.808 0.072102
finance      1.805e-01  9.122e-02   1.979 0.049140
consprod     1.688e-01  1.124e-01   1.501 0.134867
utility     -3.118e-01  9.289e-02  -3.357 0.000941
```

   The difference is 0.1805*100%=18.05%.

(b) What is the elasticity of the salary w.r.t. ROE?

Solution:

First, we take the logarithm of ROE and run a second regression:

```
logroe=log(roe)
summary_rob(lm(logsalary~sales+logroe+finance+consprod+utility))
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.594e+00  2.593e-01  25.427  < 2e-16
sales        1.332e-05  7.828e-06   1.701 0.090461
logroe       8.350e-02  8.581e-02   0.973 0.331682
finance      1.736e-01  9.233e-02   1.880 0.061501
consprod     1.937e-01  1.090e-01   1.776 0.077184
utility     -3.304e-01  9.667e-02  -3.418 0.000762
```

The elasticity is 0.0835%.

Pay attention: the ROE is, by its own nature, defined in percentage points. You could be tempted to state that the elasticity is 0.008431*100%=0.8431% looking at the ROE coefficient under point a), since you have have a log-linear model in which the regressor is already in percentage points.

However, what the coefficient under a) gives is the effect of one point more in ROE, which is not affected by its initial level (as in a simple linear regression). Elasticity instead is the effect of a change of one percentage point in ROE: ROE*(0.01), and here the initial value matters!

To see the difference, suppose the current level of ROE is 30 points, that is 30%. The coefficient under a) tells the effect on 1% more, so if ROE goes from 30 to 31%. The coefficient here estimated tells the effect of a change from 30 to 30.3% (30%*1.01).

This is why we cannot consider directly the coefficient of the regression under point a) as our elasticity.

2. The dataset *equipment.dta* contains data on transportation equipment manufacturing for 25 US States. It contains 25 observations on the following variables:

- *valueadded*: aggregate output in millions of 1957 dollars;
- *capital*: capital input in millions of 1957 dollars;
- *labor*: aggregate labor input, in millions of worked hours.

Standard models in macroeconomics, like the Solow model, assume Constant Return to Scale for the production function. Assume that the production function of the transport sector in 1957 was a Cobb-Douglas, i.e.:

$$Y = AK^{\alpha}L^{\beta}$$

(a) Estimate α and β and interpret them.

Solution:

First, apply the logarithmic transformation:

$$log(Y_i) = log(AK_i^{\alpha}L_i^{\beta} = log(A) + \alpha log(K_i) + \beta log(L_i)$$

so that you estimate:

$$log(valueadded_i) = \beta_0 + \beta_1 log(capital_i) + \beta_2 log(labor_i)$$

```
logvalue=log(valueadded)
logcapital=log(capital)
loglabor=log(labor)
summary_rob(lm(logvalue~logcapital+loglabor))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.8444	0.3102	5.947	5.52e-06
logcapital	0.2454	0.1029	2.384	0.0262
loglabor	0.8052	0.1407	5.723	9.33e-06

Thus, α is 0.24: increasing capital by 1

(b) Test the Constant Return to Scale assumption (i.e. $\alpha + \beta = 1$) [Hint: redefine the dependent variable].

Solution:

A little trick is necessary here. The aim is to test $\alpha + \beta - 1 = 0$.
Start from the former regression:

$$log(Y_i) = \beta_{inter} + \alpha log(K_i) + \beta log(L_i) + u_i$$

Now add and subtract $\beta log(K)$ from the left-hand side of the equation and subtract to both sides $log(K)$.

$$log(Y_i) - log(K_i) = \beta_{interc} + \alpha log(K_i) + \beta log(L_i) + \beta log(K_i) - \beta log(K_i) - log(K_i) + u_i$$

Rearranging the terms we get:

$$log(Y_i) - log(K_i) = \beta_{interc} + (\alpha + \beta - 1)log(K_i) + \beta(log(L_i) - log(K_i)) + u_i$$

Hence, define two new variables and run the regression:

```
xnew=loglabor-logcapital
ynew=logvalue-logcapital
summary_rob(lm(ynew~logcapital+xnew))
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.84442    0.31015   5.947 5.52e-06
logcapital   0.05061    0.04839   1.046    0.307
xnew         0.80518    0.14069   5.723 9.33e-06
```

What we are interested in is the coefficient for $log(K)$. As you see, it is not statistically different from 0. We cannot reject the null hypothesis, that is, the Constant Return to Scale assumption.

3. The dataset *wage1.dta* contains 526 observations on the following variables:

- *wage*: monthly earnings;
- *educ*: years of education;
- *female*: dummy variable, 0 if man, 1 if female.

Suppose that we would like to test whether the percentage return to education is the same for men and women, allowing for a constant wage differential between men and women (a differential for which we have already found evidence). For simplicity, we include only education and gender in the model.

(a) What kind of model allows for a constant wage differential as well as for different returns to education?

Solution:

Since we have been asked for a percentage return, we need to build a log-linear model. To test whether there is a different return to education for men and women, we need to include in the regression an interaction term, *female*educ*:

$$log(wage_i) = \beta_0 + \beta_1 female_i + \beta_2 educ_i + \beta_3 (female_i * educ_i) + u_i$$

For men, the model becomes:

$$log(wage_i) = \beta_0 + \beta_2 educ_i + u_i$$

while for women we can rewrite the model in this way:

$$log(wage_i) = \beta_0 + \beta_1 + (\beta_2 + \beta_3) educ_i + u$$

Therefore β_1 is the difference between the intercepts for men and women, while β_3 is the difference in the slope for the education variable.

```
logwage=log(wage)
summary_rob(lm(logwage~educ+female+educ:female))

               Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.260e-01  1.178e-01   7.009 7.45e-12
educ         7.723e-02  9.067e-03   8.517  < 2e-16
female      -3.601e-01  2.126e-01  -1.694   0.0909
educ:female -6.411e-05  1.698e-02  -0.004   0.9970

Residual standard error: 0.4459 on 522 degrees of freedom
Multiple R-squared: 0.3002, Adjusted R-squared: 0.2962
F-statistic: 74.65 on 3 and 522 DF,  p-value: < 2.2e-16
```

Variable1:variable2, in this case educ:female, allows for the interaction of two orthogonal factors. The wage for women is on average 36% less than for men, ceteris paribus. As you may notice, the coefficient is not significant at 5% level - even if 'by a whisker'. It is in fact significant at a 10% level, for example. Every year of education increases wage by 7.723% for men, and of $\beta_1 + \beta_3 = 7.723 - 0.0064 \simeq 7.7166\%$ for women. The coefficient is statistically significant. For the β_3 interpretation, look at next question.

(b) Test that $\beta_3 = 0$ at 5% level. What are you testing?

Solution:

We are testing whether there are differences in return to education for men and women; in other terms, whether two individuals with the same level of education will receive different salary for this characteristic only because they belong to different genders. $beta_3$ can also be seen as the difference in the slopes for education between men and women: $\frac{\delta lwage}{\delta educ}$. The t-value is smaller than 1.96 and the p-value is higher than 5%: hence, we cannot say that there are differences in return to education.