Logarithm transformations and interaction

April 22, 2013

1. The dataset we will use is CEOSAL1. It contains 209 observation for the year 1990 on the the following data:

   - *salary*: 1990 salary , thousands $;
   - *sales* : 1990 firm sales, millions $.

   We will not use the other variables.

   (a) Your employer wants to know what happens to the salary of CEO if the sales increases of 1%.

   (b) Your boss wants to know how the percentage salary of CEO will change if the sales increases of 1 million.

2. We will use dataset WAGE2 that contains 526 observations on the following variables (among others):

   - *wage*: Monthly earnings;
   - *educ*: Years of education;
   - $female$: Dummy variable:
     - =1 if female;
     - =0 otherwise.

   Suppose that we wish 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. 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 different returns to education?

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

   (c) Extend the model including other variables, justifying the choice of the transformations. Comment the results.

3. For this exercise we are going to use the dataset Openness.dta that contains inflation and import share data from 1973.The dataset contains all non-centrally planned countries with flexible and fixed exchange rate listed by Summers and Heston for whom data on inflation and openness are available. Inflation is measured as average annual change in the log GDP or GNP deflator since 1973. Openness is measured as the average share of imports in GDP or GNP over the years beginning in1973. This dataset contains:

   - *open*: imports as % GDP, '73-'80
   - inf: avg. annual inflation, '73-
   - *oil*: dummy variable :
     - =1if major oil producer,
     - =0 otherwise;

   This dataset have been used by Romer in order to test if countries more open have lower level of inflation. The data are collecting from 1973 for a specific reason : before the '73, there was the Bretton Woods [1] system that limited the possibility to pursue independent monetary policy.

   You will not use other variables hence you don't need the description.

   (a) Construct a statistical procedure to test whether oil countries and non-oil countries have the same level of inflation, against the Alternative Hypothesis that they have different level.

   (b) Regress inf on *oil*. How can you interpret β_1? And β_0? Do you notice some similarities with the point (a)?

[1] The chief features of the Bretton Woods system were an obligation for each country to adopt a monetary policy that maintained the exchange rate of its currency within a fixed valueÑplus or minus one percentÑin terms of gold and the ability of the International Monetary Funds to bridge temporary imbalances of payments. Then, on August 15, 1971 the United States unilaterally terminated convertibility of the dollar to gold. This action created the situation whereby the United States dollar became the sole backing of currencies and a reserve currency for the member states. In the face of increasing financial strain, the system collapsed in 1971.

4. For this exercise we use Crime.dta file that contains data on arrests during the year 1986 and other information on 2725 men born in either 1960 or 1961 in California. Each man in the sample was arrested at least once prior to 1986. The variables of interest are:

   - *narr*86: number of times the man was arrested during 1986, it is zero for most of the sample (72.29%) and it varies from 0 to 12;
   - *pcnv*: is the proportion of arrests prior to 1986 that lead to conviction;
   - *avgsen*: it is the average sentence length served for prior convictions;
   - *ptime*86 is the months spent in prison in 1986;
   - *qemp*86: is the number of quarters during which the man was employed in 1986 (from zero to four).

   pcnv should be interpreted as a proxy for the likelihood to be convicted of a crime while *avgsen* is a measure of expected severity of punishment, if convicted. *ptime*86 captures the incarcerative effects of crime: if an individual is in prison, he cannot be arrested for a crime outside the prison. Labor market opportunities are crudely captured by *qemp*86.

   (a) Regress *narr*86 on the other variables except *avgsen*. Interpret the coefficients and their significance.

   (b) Now do the same regression of point (a) adding *avgsen*. Does this new variable help you in explaining *narr*86? Why?

   (c) Do you think that there is a causal effect between *narr*86 and *pcnv*[2]?

[2]Causal effect means that $E(pcnv, u) = 0$. There is no variables that has an effect on the proportion of arrests prior to 1986 and contemporary on *narr*86 except the variables we put in the regression