## Exercises Econometrics: Set 1

Submit by 8:00am, Friday 29th September 2023

General guidelines:

Make sure to be able to carefully <u>explain</u> your answers and <u>interpret</u> the results. The objective of this class is not only to understand and perform data analysis using different econometric techniques, but also to help you interpret the results of empirical studies. It is the interpretation that is often the most interesting part of empirical work, in particular if we want to derive policy implications from our findings.

Exercises:

A) The data set SLEEP75.dta from Biddle and Hamermesh (1990) contains data on the number of minutes per week spend sleeping (*sleep*) and the number of minutes spend in paid work (*totwrk*).[1]

When the following simple regression was estimated using data on n = 706 individuals
$$sleep = \beta_0 + \beta_1 totwrk + u$$

the estimated slope coefficient was –0.15, while the estimated intercept was 3586.38.

1) Interpret the slope coefficient.
2) What is the predicted number of minutes of sleep for somebody who does not work in a paid job the whole week? Comment.
3) To predict a sleeping time of 8 hours a night, what would *totwrk* have to be?
4) What kinds of factors are contained in u? Are these likely to be correlated with the amount of time spend working?
5) Will a simple regression analysis uncover the ceteris paribus effect of the time spend working on the time spend on sleeping? Explain.

B) Imagine you work in a poor country where many children drop out of school, or are often absent. You are interested in knowing the impact of a program providing free school meals on learning outcomes. As you are worried about omitted variable bias, you randomly provide school meals in some schools, and not in others. When you obtain the results of the end-of-the-year exam you find that schools with free meals have lower average test results?
  1) What might be the reason for this finding?
  2) Should you use these results to advise against free school meals?
  3) What type of other data may be useful to collect in order to determine the impact of the program on children's learning?

---

[1] J.E. Biddle and D.S. Hamermesh (1990), "Sleep and the Allocation of Time," *Journal of Political Economy* 98, 922-943.

Computer assignment:

The data set busind.dta contains information on Gross National Income (GNI) per capita and the number of days to open a business and to enforce a contract in a sample of 155 countries. It was extracted from the "Doing Business" dataset, a dataset collected by the World Bank based on expert opinions in each country.[2] The variable *gnipc* measures GNI per capita in $. The variable *daysopen* measures the average number of days needed to open a business in that country, and *daysenforce* measures the average number of days needed to enforce a given type of contract.

(i)     Find the average GNI per capita and the average number of days to open a business, and the average number of days to enforce a contract.

(ii)    In how many countries does it take on average less than 5 days to open a business? What is the maximum number of days to open a business in the dataset? In which countries does it take more than 200 days to open a business?

(iii)   Estimate the following simple regression model

$$gnipc = \beta_0 + \beta_1 daysopen + u \,.$$

Give a careful interpretation of estimate of $\beta_1$ and $\beta_0$. Are the signs what you expected them to be?

(iv)    What kinds of factors are contained in u? Are these likely to be correlated with the number of days to open a business ?

(v)     What is according to this model the predicted income for a country where it takes 5 days to open a business? And the predicted income for a country where it takes 200 days to open a business? Show how you can calculate the answers by hand (once you have obtained the estimation results). Do the obtained levels of income seem reasonable? Explain.

(vi)    Estimate the following simple regression model and give a careful interpretation of $\beta_1$.
$$gnipc = \beta_0 + \beta_1 daysenforce + u \,.$$

(vii)   Comparing the estimates of the models in (iii) and (v), which one explains more of the variation in income per capita across countries. Can you infer whether the duration to open a business or the duration for enforcing contracts is more strongly correlated with income per capita?

(viii)  Estimate the following simple regression model:

$$\log(gnipc) = \beta_0 + \beta_1 daysopen + u$$

Give a careful economic interpretation of the estimate of $\beta_1$.

(ix)    Do these results allow you to draw conclusions regarding the desirability of policies aimed at reducing the number of days for opening a business in certain developing countries?[3]

---

[2] See World Bank website on the "Doing Business Indicators" (http://www.doingbusiness.org/) for more information on the methodology. You don't need this additional info for the problem set, the link is only provided for those of you who are curious.

[3] As you probably know, such policies are advocated by many bilateral and multilateral aid organizations.

(x)     The dataset contains 155 countries, and hence does not contain information about all the countries in the world.  Do you think one should account for that when interpreting the regression results. Why?