Econometrics problem set I.

Submission deadline: 4th October 2025, 23.59 via Moodle

Please submit via Moodle

- one *well-edited* word or pdf file that contains the (edited) solutions (it should be understandable on its own, without referring to the R script),
- one R script that contains the calculations,

Use R for all calculations.

The attached health_small.xlsx file contains the disability adjusted life expectancy (DALE, in years), the GDP per capita (GDPC, in USD) and the Gini-coefficient (GINI) for various countries for years 1993-1997. The Gini-coefficient measures the inequality of income within a country (0 being the most equal, 1 the most unequal society).

Note: You can import the dataset to R as follows. (Of course, you can use other commands as well.)

```
library(readxl)
health <- read_excel(".../health_small.xlsx")
attach(health)
```

where you should indicate the place of the file on your desktop. Note that you should use "/" in the path of the file. Afterwards, you may want to use e.g. the command attach(health) to make the variables directly accessible, or you can refer to the variables as e.g. health$DALE.

Use only one year in your analysis determined by the first character in your Neptun code: 1997 (A-C), 1996 (D-F), 1995 (G-I), 1994 (J-P), 1993 (Q-Z or number). Please check carefully that you solve the problem allocated to you.

1.

a) Plot DALE as a function of GDPC and (in a separate plot) DALE as a function of the logarithm of GDPC. Interpret the plots.
b) Estimate with OLS a model where DALE is the dependent variable and the logarithm of GDPC is the explanatory variable (besides, of course, the constant).
c) Draw the scatterplot of the residuals and the logarithm of GDPC. What do you think: are the residuals homoscedastic or heteroscedastic?
d) Calculate the usual and the heteroscedasticity-robust standard errors of the estimated parameters of the above model estimated with OLS.

In this case it is not trivial which type of standard error should be used, even if the model is heteroscedastic, because the sample size is not sufficiently large for the robust standard errors to be used with great confidence. So, in the following, use the more "conservative", i.e. the larger standard errors.

e) Interpret the coefficient of the logarithm of GDPC with your own words. Is the magnitude of the coefficient practically (i.e., from an economic point of view) meaningful?

f) What do you think: does the estimated coefficient for log(GDP) give an unbiased estimate of the causal effect of log(GDP) on DALE? Why or why not? If you argue that the estimate is not unbiased please indicate whether it is biased upwards or downwards. Please provide a detailed answer.

g) Give the 95% confidence interval of the parameter estimate of log(GDPC).

h) Test whether it is significant at the 1% level. Give the p-value of the test statistic.

i) On the top of the logarithmic plot in part a), plot the estimated regression functions of DALE as a function of logarithmic GDPC.

2.

a) Estimate a model of DALE where the quadratic function of GDPC (not the logarithmic GDPC) is used as explanatory variable.

b) Based on the latter model, calculate the partial (marginal) effect of GDPC if GDPC takes a particular value, **depending on the first letter of your surname.** It should be GDPC=6,000 USD (for surnames starting with Q-Z), GDPC=9,000 USD (with G-P) and GDPC=14,000 USD (with A-F).

c) How can this partial effect be compared to the result obtained in part 1?

3.

a) Now estimate a model with DALE as the dependent variable, and log(GDPC) and GINI as explanatory variables.

b) Interpret the parameter estimate of GINI.

c) Can we conclude that a larger GINI causes a smaller disability adjusted life expectancy? Why or why not?

d) Interpret the difference between the estimated parameter of log(GDPC) in part 1 and in part 3.

e) Illustrate the omitted variable formula in this case.