Data Based Economics: Course Work

1. **A visualization exercise**

Choose a simple graph from https://ourworldindata.org/ and try to reproduce it as well as you can (don't forget to provide a link to the original). You can use the data that is usually attached with the graphs. It is ok to simplify the graphs features (e.g. no interactivity for instance).

1. **Feldstein-Horioka puzzle**

In a famous paper, *Domestic saving and international capital flows* (Economic Journal, 1980), Martin Feldstein and Charles Horioka, exposed the following puzzle: - if international flows of capital were frictionless, capital allocations should be determined by productivity differentials - as a result, fluctuations in investment in a given country should be primarly explained by capital flows - yet, the data shows a strong correlation between *domestic* saving and domestic investment

Your goal is to confirm the last point of the Feldstein-Horioka analysis by estimating the following regression across OECD countries:[1]

$$\frac{I_{it}}{Y_{it}} = \alpha + \beta \frac{S_{it}}{Y_{it}} + \epsilon_{it}$$

where I_{it}, S_{it} and Y_{it} are gross investement, saving and total production in country i at date t and ϵ_{it} is a gaussian random error.[2]

In this regression a low coefficient of β means a high mobility of capital (local investment is determined by productivity not local savings), while β close to one corresponds to high capital mobility.

We would like to estimate this relationship for as many OECD countries as possible from 1970 to 2010 and to check the results on subperiods 1970-1979, 1980-1989, 1990-1999 and 2000, 2009.

I suggest the following steps:

a. Download national accounts data for as many OECD countries as possible (from OECD website or dbnomics) with the widest time period interval (even if some countries don't report results for particular years).
b. Construct a panel with all data in one dataframe
   - four columns: date, country, investment, saving, gdp
c. Describe the dataframe, summarize data availability, make some relevant plots. . .

---

[1] Given the accounting relation $I = S - (X - M)$ the regression can equivalently be rewritten as $\frac{-(X_i - M_i)}{Y_i} = \alpha + \beta \frac{S_i}{Y_i} + \epsilon$. In other words it tests whether net capital flows $X - M$ can be explained by fluctuations in domestic savings.

[2] Note that this specification is a vanilla OLS regression. α and β are not dependent on the country or on time.

d. Compute the investment over gdp and saving over gdp ratios.
e. For each subperiod:
- select the relevant subpart of the database
- run the regression
- check its validity
- comment the results briefly
f. (Freestyle) Discuss the results, propose an extension.