

Homework 9

Statistical Computing, 36-350

Due Wednesday Nov 18, 2015

Gross domestic product (GDP) is a measure of the total market value of all goods and services produced in a given country in a given year. The percentage **growth rate of GDP** in year t is

$$100 \times \left(\frac{GDP_{t+1} - GDP_t}{GDP_t} \right) - 100$$

An important claim in economics is that the rate of GDP growth is closely related to the level of government debt, specifically with the ratio of the government's debt to the GDP. The file <http://stat.cmu.edu/~ryantibs/statcomp/labs/debt.csv> on the class website contains measurements of the GDP growth rate (column name **growth**) and of the debt-to-GDP ratio (column name **ratio**) for twenty countries around the world, from the 1940s to 2010. Note that not every country has data for the same years, and some years in the middle of the period are missing data for some countries but not others.

(This data is also used in Lab 10.)

1. Load the data into a data frame named **debt** and make a scatter-plot of the GDP growth rate (vertical axis) against the debt ratio (horizontal axis).
2. Use **daply** to compute the mean growth rate and debt ratio for each **year** in the data set. Plot the results.
3. Fit a linear model of growth on the debt ratio, using **lm()**. Report the intercept and slope to reasonable precision. (Explain why the precision you give is reasonable.) Add a line to your scatterplot from question 1 showing the fitted regression line. Visually, is it a reasonable match to the data?
4. Some economists claim that high levels of government debt cause slower growth. Other economists claim that low economic growth leads to higher levels of government debt. The data file, as given, lets us relate this year's debt to this year's growth rate; to check these claims, we need to relate current debt to future growth.
 - a. Create a new dataframe which just contains the rows of **debt** for France, but contains all those rows. It should have 54 rows and 4 columns. Note that some years are missing from the middle of this data set.
 - b. Create a new column in your dataframe for France, **next.growth**, which gives next year's growth *if* the next year is in the data frame, or **NA** if the next year is missing. (**next.growth** for 1971 should be (rounded) 5.886, but for 1972 it should be **NA**.)
5. Add a **next.growth** column, as in question 4, to the *whole* of the **debt** data frame. Make sure that you do not accidentally put the first growth value for one country as the **next.growth** value for another. (The **next.growth** for France in 2009 should be **NA**, not 9.167.)

Hints: Write a function to encapsulate what you did in question 4, and apply it using **ddply()**.
6. Make a scatter-plot of next year's GDP growth against this year's debt ratio. Linearly regress next year's growth rate on the current year's debt ratio, and add the line to the plot. Report the intercept and slope to reasonable precision. How do they compare to the regression of the current year's growth on the current year's debt ratio?
7. Make a scatter-plot of next year's GDP growth against the current year's GDP growth. Linearly regress next year's growth on this year's growth, and add the line to the plot. Report the coefficients (to reasonable precision). Can you tell, from comparing these two simple regressions (from the current question, and question 6), whether current growth or current debt is a better predictor of future growth?

8. Add a column, `delta.growth`, to the `debt` dataframe, giving the difference between next year's GDP growth rate and this year's GDP growth rate. Then regress the change in GDP growth on the current GDP growth and the current debt level. Report the coefficients.
9. Some economists have claimed that there is a “tipping point”, or even a “point of no return” when the ratio of government debt to GDP crosses 90%, above which growth slows dramatically or even becomes negative. Add a column `high.debt` to the dataframe that is true when the debt ratio is over 90% and false when it is not. Repeat the regression from question 8, adding `high.debt` to your model as a third (“dummy”) variable. Report the coefficient of `high.debt`; what does its value tell you about the claim?

Behind the scenes: It'd be a spoiler to say where the problems came from, or even the data. With a little Googling, you can probably figure it out!