Homework 13: Over the Cliff

36-402, Advanced Data Analysis

Due at 6:00 pm on Thursday, 24 April 2025

AGENDA: Fitting models of relations of variables over time

An important and controversial question in macroeconomics and political economy is whether high levels of government debt causes the economy to grow more slowly or even shrink. There are several plausible reasons why it might[1]; some economists claim that there is a threshold level of debt, perhaps around 90% of GDP, above which growth rates plummet.

Against this, there are other reasons why high levels of debt might *not* cause growth to slow, at least not always[2]. In particular, since "high levels of government debt" are defined relative to the size of the economy, as a high ratio of debt to GDP, slow growth itself might cause higher levels of government debt.

This week's data set, `debt.csv`, contains information on GDP and government debt for a selection of countries since World War II. For each country and year, we should have the GDP (nominal, i.e., not adjusted for inflation or differences in exchange rates) and the size of government debt (also nominal). Unfortunately, one or both values may be missing for some countries in some years.

1. (6) The data set contains a variable, `growth`, which is the annual growth rate in real (inflation-adjusted) GDP for each country and year. It also contains a variable, `ratio` which is the ratio of government debt to GDP. Make a scatter-plot with `growth` on the vertical axis and `ratio` on the horizontal. Describe the patterns you see, if any.

2. (10) Run a nonparametric regression of `growth` on `ratio`, and plot the resulting curve. Describe and interpret the curve. Does it suggest an abrupt slowing of growth above some threshold level of debt?

[1] High levels of government borrowing might "crowd out" investing in the private sector, by using up available savings and/or raising the interest rates at which businesses can borrow; capitalists might anticipate that the debt will either be paid off through high taxes or discharged through inflation, and prefer to spend their money on luxuries now, rather than invest and see the investment go away later; high levels of debt might lead to lower confidence that the government generally knows what it's doing, making investment seem too risky; etc.

[2] A depressed economy has unused resources, so government borrowing which employs those resources needn't lead to crowding out; the things government spends money on (roads, schools, hospitals, basic research, honest markets) increase the value of private investments; governments which can borrow large sums are receiving a market endorsement of their willingness and ability to pay their debts; etc.

3. (6) Since changes in government debt levels might take some time to affect economic growth, we would like to compare growth in year $t+1$ to `ratio` in year t. Create a new variable, `growth.lead1`, which records for each country/year the *next* year's GDP growth, with NAs in the right places when it is not available. Add `growth.lead1` to the data frame. Plot `growth.lead1` against `growth`, and describe any patterns you see.

   Hints: Make sure that you do not confuse growth rates from different countries (so that, e.g., the last year for Austria gets a growth rate from Belgium). You may find Recipes 14.7 (and 6.6) from *The R Cookbook* helpful.

4. (6) Plot `growth.lead1` against `ratio`, and do a nonparametric regression of the former on the latter. Describe the results, and compare them to those of Q2.

5. (10) Estimate an additive model where `growth.lead1` is predicted from `growth` and `ratio`. Is the partial response to the previous year's growth nearly linear? Should it be? Compare the partial response function for debt to the curves from problems 2 and 4.

6. (6) Create a new variable, `growth.lag1`, which represents the *previous* year's growth rate (with NAs in appropriate places), and add it to the data set. Plot it against `ratio` and fit a nonparametric regression. Does `ratio` do a better job of predicting `growth` or `growth.lag1`?

7. (10) Estimate an additive model in which the current year's `ratio` is predicted by last year's `ratio`, last year's `growth`, and the current year's `growth`. (You may have to create a new column.) Describe the partial response functions, and whether any predictor variables could be dropped.

8. (15) What we would have to assume for the model in Q5 to give us an unconfounded estimate of the causal effect of government debt on future economic growth? Be as specific as possible. (You may want to draw some DAGs, and include them in your write-up.) Comment on how plausible those assumptions are, and on what might go wrong if they fail.

9. (1) *Timing* How long, roughly, did you spend on this assignment? How much of that time was spent on math, on coding/debugging, and on writing?

Presentation Rubric (15): The text is laid out cleanly, with clear divisions between problems and sub-problems. The writing itself is well-organized, free of grammatical and other mechanical errors, and easy to follow. Plots are carefully labeled, with informative and legible titles, axis labels, and (if called for) sub-titles and legends; they are placed near the text of the corresponding problem. All quantitative and mathematical claims are supported by appropriate derivations, included in the text, or calculations in code. Numerical results are reported to appropriate precision. All parts of all problems are answered

with actual coherent sentences, and raw computer code or output are only shown when explicitly asked for. Text from the homework assignment, including this rubric, is included only when relevant, not blindly copied.

(In Gradescope, assign *all* pages to this rubric.)

Code Rubric (15): The code is logically organized and easy to read. No redundant code; no needlessly repetitive code; no unused code. Variables and functions have descriptive and appropriate names. (Loop or array indices, arguments, etc., can have short, conventional names such as `i`, `x`, `df`, etc.) All functions have comments defining their purpose, their inputs, their outputs, and any dependencies on other code you wrote. Vectorization is used wherever appropriate. Allowed packages: `knitr`, `tidyverse`, `dplyr`, `ggplot2`, and those explicitly mentioned in the textbook or the assignment for implementing particular methods. (Any other packages require prior permission from the professor, which must be renewed for each assignment; record the date on which you got permission in your comments.) Code from the textbook and class examples is used wherever possible and appropriate. In particular, it should be used for tasks like bootstrapping, calibration plots, and cross-validation (*unless* the package implementing a model includes its own cross-validation functions). All plots and tables are generated by code included in the R Markdown file. Numerical results (etc.) appearing in text are neither hand-copied nor spat out with `cat()`, `print()`, `sprintf()` etc., but instead properly formatted through in-line code.

(Do not assign any pages to this rubric; instead, submit your Rmd file to the "HW k R Markdown File" assignment on Gradescope, for the appropriate k.)