Homework 2: …But We Make It Up in Volume

36-402, Spring 2025

Due at 6 pm on Thursday, 30 January 2025

"Gross domestic product" is a standard measure of the size of an economy; it's the total value of all goods and services bought and solid in a country over the course of a year. It's not a perfect measure of prosperity[1], but it is a very common one, and many important questions in economics turn on what leads GDP to grow faster or slower.

One common idea is that poorer economies, those with lower initial GDPs, should grower faster than richer ones. The reasoning behind this "catching up" is that poor economies can copy technologies and procedures from richer ones, but already-developed countries can only grow as technology advances. A second, separate idea is that countries can boost their growth rate by under-valuing their currency, making the goods and services they export cheaper for purchasers in other countries.

This week's data set contains the following variables:

- Country, in a three-letter code (see `http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3`).
- Year (in five-year increments).
- Per-capita GDP, in dollars per person per year ("real" or inflation-adjusted).
- Average percentage growth rate in GDP over the next five years.
- An index of currency under-valuation[2]. The index is `0` if the currency is neither over- nor under- valued, positive if under-valued, negative if it is over-valued.

Note that not all countries have data for all years. However, there are no missing values in the data table.

[1] A standard example: if vandals break all the windows on a street, GDP goes *up* by the cost of the repairs.

[2] The idea is to compare the actual exchange rate with the US dollar to what's implied by the prices of internationally traded goods in that country — the exchange rate which would ensure "purchasing power parity". The details are in the paper this assignment is based on, which will be revealed in the solutions.

When you submit your work through Gradescope, be sure to assign pages to problems. You will lose points if you do not.

1. (6) Chapter 3, exercise 3, part 1. *Hint:* The in-class exercise for 21 January.
2. (5) Linearly regress the growth rate on the under-valuation index and the log of GDP. Report the coefficients and their standard errors. Do the coefficients support the idea of "catching up"? Do they support the idea that under-valuing a currency boosts economic growth?
3. Repeat the linear regression but add as covariates the country, and the year. Use `factor(year)`, not `year`, in the regression formula.
   (a) (5) Report the coefficients for log GDP and undervaluation, and their standard errors.
   (b) (5) Explain why it is more appropriate to use `factor(year)` in the formula than just `year`.
   (c) (5) Plot the year coefficients as a function of time.
   (d) (5) Does this expanded model support the idea of catching up? Of undervaluation boosting growth?
4. Does adding in year and country as covariates improve the predictive ability of a linear model which includes log GDP and under-valuation?
   (a) (2) What are the R^2 and the adjusted R^2 of the two models?
   (b) (5) Use leave-one-out cross-validation to find the mean squared errors of the two models. Which one actually predicts better, and by how much? *Hint:* Use the code from lecture 3 (and/or chapter 3 of the textbook).
   (c) (5) Explain why using 5-fold cross-validation would be hard here. (You don't need to figure out how to do it.)
5. *Kernel smoothing* Use kernel regression, as implemented in the `np` package, to non-parametrically regress growth on log GDP, under-valuation, country and year. *Hint:* read chapter four carefully. In particular, try setting `tol` to about 10^{-3} and `ftol` to about 10^{-4} in the `npreg` command, and allow several minutes for it to run. (You'll probably want to cache this part of your code.)
   (a) (5) Give the coefficients of the kernel regression, or explain why you can't.
   (b) (5) Plot the predicted values of the kernel regression, for each country/year combination, against the observed values of GDP growth.
   (c) (5) Plot the residuals of the kernel regression against its predicted values. Should these points be scattered around a flat line, if the model is right? Are they?
   (d) (5) The `npreg` function reports a cross-validated estimate of the mean squared error for the model it fits. What is that? Does the kernel regression predict better or worse than the linear model with the same variables?

6. *Time courses and scenarios* In this question, use the kernel regression you fit in the previous problem. (You can combine plots from parts (a)–(e) if that's helpful, and you clearly indicate which parts you are combining.)

   (a) (5) Plot the predicted growth rate, as a function of the year, in five year increments from 1955 to 2000, if the initial GDP (not log GDP!) is $10,000 in each period, the under-valuation index is 0 (i.e., no under- or over- valuation), and the country is Brazil .

   (b) (4) Re-do the plot but change the under-valuation index to +0.4.

   (c) (4) Re-do the plot but hold the initial GDP at $3,20 and the under-valuation index at 0.

   (d) (4) Re-do the plot with the initial GDP at $3,200 and the under-valuation index at +0.4.

   (e) (4) Re-do the plot from Q6a but change the country to Thailand.

   (f) (5) For the same kernel regression, plot the predictions for growth against each variable, holding the other variables fixed at their medians, or, for country, the mode. (You can get these plots using the `plot` command and the kernel regression object from `npreg`.) Is there evidence that $\log(GDP)$ or under-valuation are strongly related to growth? Which variables do show a strong relationship with growth?

7. (1) How long, roughly, did you spend on this problem set?

PRESENTATION RUBRIC (10): 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 plots and tables are generated by code included in the R Markdown file. 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.)