Project 2

Quantitative Methods (Business Analytics and Supply Chain) — Fall 2018

Due: Tuesday, October 16 at 5pm

This project asks for you to collect and work with data in a realistic setting. For this assignment,

1. Write a **brief and insightful typed report** on your findings using the data.
2. As an appendix to your report, include the relevant R code and output (script files, log files, etc.) you used to perform the calculations you discuss in the report.
3. In the main text of your report, you should have typed tables and figures that help you summarize nicely the **relevant** computations and graphics from this output (not all output is relevant to a coherent discussion).

Be careful to present your work with proper grammar, and professional tone. Your grade will reflect not only your calculations, but the manner in which you interpret them and the form in which you present them. You should submit your final project write up within Canvas (together with R script and log files) by the due date.

John Taylor's Other Scatter Plot

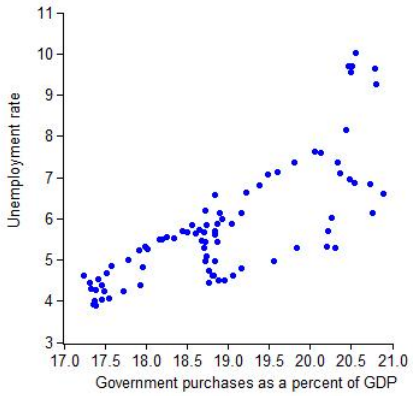
In 2011, an economist named Justin Wolfers wrote a blog post about a striking scatterplot produced by another economist John Taylor on his own blog. Read Taylor's original post,[1] Wolfers' first response,[2] and Wolfers' second response.[3] In the infamous post that generated the controversy, John Taylor actually had two scatter plots. Although the side discussion is interesting controversy, this exercise asks you to produce the other scatter plot – one that got considerably less attention, but is nonetheless interesting. Here is the other scatter plot and how Taylor describes it in his post:

[1] https://economicsone.com/2011/01/14/higher-investment-best-way-to-reduce-unemployment-recent-experience-shows/

[2] http://www.freakonomics.com/2011/03/30/how-to-spot-advocacy-science-john-taylor-edition/

[3] http://freakonomics.com/2011/04/01/graph-fight-more-on-taylors-scatterplot/

Some economists argue that the efforts now underway to reduce government spending as a share of GDP will have adverse effects on unemployment. This is not what the data show. Consider this chart which shows the pattern of government purchases as a share of GDP and the unemployment rate over the past two decades. (The data are quarterly seasonally adjusted from 1990Q1 to 2010Q3.) There is no indication that lower government purchases increase unemployment; in fact we see the opposite, and a time-series regression analysis to detect timing shows that the correlation is not due to any reverse causation from high unemployment to more government purchases.



Your task is as follows:

1. Construct the data set that generates this scatter plot. The source is FRED, [4] a great source for data on macroeconomics aggregates. When you download the data from FRED, it is helpful to save in .csv format and read into R using read.csv(). Make sure you have obtained the right series of data. You may need to transform data you can download into the variables that Taylor plots. It may be helpful to merge the data into one data frame (merge() is useful for this). Finally, for the later parts of this problem, make sure you download the entire series of observations, not just the ones that Taylor plots.

2. Use the data set you constructed to reproduce Taylor's other scatter plot. That is, produce a scatter plot using only the observations from Q1 of 1990 to Q3 of 2010.

   (a) As you do this, try to format the dates as dates (look up how to do this using the as.Date() function). This will help you to produce nice looking plots within R.

3. Compare the scatter plot you obtain using the full data set to the one that Taylor reports.

   (a) For this comparison, you may want to use different plotting characters and colors, and the points() function to produce different plotting characters for different parts of the sample.

4. Analyze the statistical relationship between seasonally-adjusted unemployment and government expenditures in a way that is informative and insightful. Some questions you should consider as you perform the analysis:

[4] http://research.stlouisfed.org/fred2/

(a) Is the correlation that Taylor presents in his scatter plot stable over time? In your analysis, you could compute the correlation separately decade. In addition to calculations, your analysis should include informative plots that support your main points.

(b) Write up your comments *concisely* in a typed report.

5. Write code – likely using for loops – to simulate the process of "cherry picking" in Wolfers' terminology, and use it to evaluate the seriousness of this part of the criticism. Some suggestions and refinements to guide your coding journey.

   (a) The code should generate two uncorrelated variables x and y (e.g., from a Normal distribution) over a time series of 256 quarters

   (b) Using these variables, use R to compute the correlation between x and y over the last 60 quarters, over the last 61 quarters, over the last 62 quarters.... all the way until... over the last 256 quarters (the full sample).

      i. A researcher who is cherry picking the begin date will then select the lowest correlation. Have your code do this and store the maximum "cherry picked" correlation, but for comparison, you should also store the correlation over the entire sample ("not cherry picked").

   (c) Repeat this process for 1000 possible "cherry picked" correlations and "not cherry picked" correlations. Store the output.

      i. Are the cherry picked correlations different than the not cherry picked correlations? Use both visual (plots to compare) and numerical evidence (t-tests).

      ii. Using the simulated cherry picked sample of correlations, conduct a t-test to evaluate whether Taylor's computed negative correlation could be generated from this cherry picking process with zero underlying correlation.