Homework 4

Gather data where there is a numerical response variable, and at least three numerical potential predicting variables (not including transformed versions of the response or other predictors), and the cases are ordered in time (e.g., divorce rates from marriage rates, birth rates, percentage of the female population that have a college degree, and per capita income for New York during the years 1940–2017; national unemployment rate from prime interest rate, return on the S&P 500, change in the U.S. dollar/Euro exchange rate, and inflation rate for quarters in 1974–2018; etc.). You can use time itself as a predictor if you wish, but there should be at least three (potential) predictors in your model other than time. There should be at least 30 data points (that is, 30 consecutive time points at which data were measured). Don’t forget that your observations must be ordered forwards in time (from least to most recent), not backwards in time. Perform a complete and full analysis of the regression model, being careful to check all appropriate assumptions and provide all relevant discussion. Use corrective procedures where necessary. Discuss the implications of your results. If it turns out that there’s no autocorrelation in your data, that’s fine — you’ve been lucky (this result might very well be surprising enough to warrant specific comment and discussion, in fact). On the other hand, if there is autocorrelation present, you must make some attempt at addressing it. As always, your goal is to find the “best” model you can to describe the underlying random process, whether than means simplifying the model, omitting unusual observations, using transformed variables, or anything else that is appropriate.

Note, by the way, that values taken at a set of time points for a sample of objects (for example, annual data for a sample of 50 countries for 2011-2017) are not time series data, and should not be used here; those are what are called panel data, and must be analyzed using (other) appropriate methods.

I’m only expecting you to implement the corrective actions for autocorrelation that we’ve discussed in class. If after you’ve done that autocorrelation (or some other model violation that we have not discussed how to address) is still present, then that’s what you
should report. In that case you should still discuss the implications of your “best” model, while also (of course) noting the potential limitations of those implications. That includes using the methods for choosing among a set of potential models that we have discussed in class, while also recognizing and discussing their potential for providing incorrect or misleading descriptions of and inferences about your data.

If you used time series data for Homework 3, you can build on that analysis here (using additional or other variables, for example). You should **not** use the same underlying data for this assignment as was used there, however. Also, a little hint: economic and financial data are particularly likely to be in time series form (and you shouldn’t be surprised if 2008-2009 cause trouble). The entries under “Links to Economic and Financial Data” in the “Links to Useful Data” page have links to literally thousands of time series data sets, at daily, weekly, monthly, or annual frequencies (depending on the series). See http://www.stern.nyu.edu/~jsimonof/classes/datalink.html

A reminder: get your data from an original data source. As was stated in the syllabus, you should **not** take your data from a textbook, a journal article that includes (time series) regression analysis of the data, an online digest of data sets that have been put together for teaching or expository purposes, or a data analysis competition (see the previous two homeworks for lists of the kinds of sites this refers to). This also includes digests of data sets from textbooks and articles specifically gathered together to be examples of time series regression (that is, **don’t** try to find data by doing a Google search of “time series regression data” or use data from a web page with the title “Time series regression data sets”).

Please be sure to include a cover page for your homework that has your name on it. It needn’t have anything else on it, but it should definitely not contain any of the text of your homework.

**Due date: April 18**