DSC 530 Final Project Write Up

Aaron Dighton

The outcome of my exploratory data analysis is that some of my variables are significant, while others may not have been good choices for my final regression. Also, variables that I picked because I thought they would be highly correlated (and need to be corrected) were not that correlated and I may need to correct for time lags between effects in the economy.

I may need to look at special cases that caused outliers in the data (I assume COVID is responsible for the incredibly high values in M1, for example, that may need to be removed to get an accurate representation of how the money supply affects GDP growth). There is also the high level of unemployment that may need to be checked separately from the other data.

Some additional variables I would like to test are the time lagged version (and maybe time advanced versions) of certain variables. I believe changes in the money supply and unemployment would predate changes in GDP, while changes in inflation would come after the fact (that might make it not a good variable for predicting GDP, however). I could also add net exports in, I did not expect government expenditures to have such a large, statistically significant effect, and I was originally using it as a kind of placeholder. Perhaps net exports would have a compounded effect on GDP as well.

I assumed some of these effects would happen in the same time period, which is obviously not correct. What did surprise me was that some of the variables were statistically significant over the 80 year data set given this incorrect assumption. Government expenditures makes sense, as government expenditures are somewhat set, and grow, each year, along with GDP. The M1 money supply also makes sense as the Federal Reserve is probably responding to economic conditions. However, the unemployment rate shocked me that current unemployment rate changes have such a drastic relationship with concurrent changes in GDP.

Challenges I faced included not knowing which time-varied variables to include. Some research I should do is whether there are known delays between certain actions and their effects on the economy. I also noticed some strange effects when I was trying to create bins for the PMF comparison. I’m not sure what was happening but if I made extra bins (beyond 1 for each % of inflation up to 13-14%), the resulting chart would be able to be extended past the levels of inflation that were in the data set. I assume this is some problem with the bin function that I wasn’t understanding.