Econ 313, Fall 2020

HW #2

For this assignment, you will forecast the unemployment rate for a country of your choosing (you may pick the U.S.) over the next two years. In doing so, you will employ a vector autoregression (VAR), a commonly used technique in applied macroeconomic work.
This is a challenging assignment. I will show examples of this method in class during the week of September 14-18.

Here are some class notes some idiot wrote for VARs. There are plenty of free guides to VARs online, although most of them require familiarity with linear algebra.

Your first task is to collect a dataset. Your should include the unemployment rate and two or three other variables that you think might help explain and forecast unemployment. It should also include a variable measuring time.

1. Explain your choice of variables and provide descriptive statistics for each of them. You may present your statistics as a table or asset of graphs.

If your variables include a trend (like GDP or stock prices), then you will want to use the percentage change for these variables instead of the levels. Otherwise, your regression results will be biased.

2. Did you choose to use the percentage change for any of your variables instead of the level? Explain.

Before you start to run any regressions, you will need to tell Stata which variable indicates time. Suppose that "date" does so. You will need to use the command "tsset date."

A VAR starts with a set of OLS regressions, one for each variable in your model. With a VAR, there is not a single dependent variable, rather there is a set of them. Unemployment will be, for example, regressed on lags of all the variables in your model.

The next step is to decide how many lags your model should have. Suppose that your system is unemployment and inflation. a VAR(1) will regress UE_t on UE_{t-1} π_{t-1}, and will also regress π_t on UE_{t-1} π_{t-1}. A VAR(2) will also include UE_{t-2} and π_{t-2} as regressions in both regressions. Note that a VAR(2) includes both the first and second lags, not just the second lag.Your next task is to decide how many lags to include.

In Stata, the command "varsoc Y" will help you do this where Y is the list of variables in your model (*e.g* "varsoc ue infl"). This command will run your VAR with different lags and use an information criteria to decide which is best.

3. How many lags did you decide to use in your VAR model?

Unlike OLS, the order of variables matters in a VAR. Suppose that I have unemployment (UE), inflation (infl) and interest rates (ir). I will then run a VAR with three lags using the command "var ue infl ir, lags(1/3)." When you run this command, however, make sure to use the number of lags from #3. If I choose this order, however, then I am assuming the following.

i) Random changes to inflation and interest rates cannot immediately affect unemployment. Instead, it must take at least one period for their to be an impact.

ii) Random changes to interest rates cannot affect interest rates right away, it must take at least one period for there to be an impact. Random changes to unemployment can, however, affect inflation immediately.

iii. Random changes to both unemployment and inflation can affect interest rates right away.

Put another way, the first variable listed is the most exogenous while the last is the most endogenous.

4. Pick and justify and order for your variables.

5. Run your VAR model using the "var" command."

Unlike OLS, we do not really care about the regression coefficients with a VAR. One common way to instead report results is an "impulse response function" (IRF) An "impulse" is a random change to one of the variables in your model. An IRF estimates how all of the variables in the model respond to this impulse. If the model's assumptions (including i-iii from above) are valid, then IRFS shed light on the causal relationships among the variables.

After running your VAR, you can calculate the IRFs by running the command "irf create 313var, step(24) set(313var)" By choosing the option "step(24)," you are asking Stata to calculate the results for 24 months ahead.

The "irf create" command calculates, but does not report, the results, You can report the results using "irf graph oirf, byopts(yrescale)" Note that switching from "graph" to "table" will instead provide a table with the results.

6. Report your IRFs.

Statistical significance can be obtained by checking whether the confidence bands include the x-axis. The IRF can be significant for some periods, but not others.

7. Describe the causal relationships suggested by your IRFs.

To compute a forecast, we will plug in the current values of each variable and use the results from #5 to estimate the most likely path of each variable. You can easily obtain the is forecast using the command "fcast compute fut, step(24)" Here "fut" is a prefix that will be applied to teh forecasted value for each variable. "Step(24)" specifies that you want Stata to compute the forecast 24 months ahead.The forecast appears as a new variable, along with upper and lower bounds, in your dataset.

8. Report your forecast for unemployment.

9. Covid-19 has created unusual circumstances which cannot be captured in your dataset. Discuss how the omission of covid-19 data affects the validity of your results.

10. Instead of running a VAR, you could have simply regressed unemployment on the other variables in your system using OLS. What econometric problems would have made such a model inappropriate?