TAKEHOME FINAL EXERCISE

As we discussed in one class lecture, when the structural shocks in a SVAR have variances that differ across observations in a pattern that is known or can be estimated, we can identify the model, (but not necessarily the names of the particular shocks) with no other restrictions. In this exercise you apply this idea to a bivariate VAR relating a credit-to-GDP ratio to real GDP. The data are available across several countries, so we might expect to be able to achieve identification by using cross-country variation in structural shocks. A working hypothesis might be that there are two kinds of shocks: "bad bubble" shocks that increase the credit-to-gdp ratio and perhaps temporarily increase real GDP, but before long lead to decline in both; and "productivity news" shocks, that increase the credit-to-gdp ratio because investors borrow to take advantage of new technology that leads to sustained growth in real GDP.

The model in abstract form is

$$A(L)y_{it} = \Lambda_i \varepsilon_{it} \tag{1}$$

$$\Lambda_i \text{ diagonal }, \text{ all } i \tag{2}$$

$$\Lambda_1 = I\,, \qquad \operatorname{Var}(\varepsilon_{it}) = I, \tag{3}$$

and as usual we assume ε_{it} independent across i and t. A normalization is needed to avoid parameter redundancy, which is the reason for setting $\Lambda_1 = I$.

This enforces the same dynamics across all countries, with all differences arising from different relative sizes of shock variances. In particular, it does not allow for differences in level across countries, which is unrealistic. To allow for differences in level, we could include a country-specific constant term vector c_i in (1). However, since this involves a lot of programming bookkeeping and this is a time-limited takehome exercise, you can just first difference the data — use the change in credit-to-GDP and the change in log real GDP. We have learned that differencing loses information, so this is bad practice, but results might still be of some interest. If you decide you have time to introduce country-specific constant terms, working in levels would be better.

The R program `SVARhtskdmdd` calculates the marginal data density for a given set of Λ_i diagonals and a given A_0. You can use it, or your own code based on it, to find the peak of that marginal data density as a function of these parameters. Then you can put a prior on them and sample from their joint posterior. A reasonable prior for the diagonal elements λ_i of Λ_i would be one-sided Cauchy, with pdf $(2/\pi)/(1+\lambda_i^2)$. This makes the prior on λ_i and $1/\lambda_i$ the same, which makes sense because if we changed the country we're normalizing on, from i to j, say, the new λ_i would be $1/\lambda_j$. But the new values of λ_k for other k's would be λ_k/λ_j, and these would not be Cauchy (though they would have Cauchy-like tails). I don't know of a nice prior that would handle the normalization while treating all the i's symmetrically in this sense.

There is also a permutation issue here. Interchanging the roles of the first and second shock by permuting the columns of $A(L)$ and the rows and columns of the covariance matrices leaves the implications of the model for data behavior the same. In constructing impulse responses from sampled parameters you will need to follow a rule for assigning shocks — for example, picking as the first shock the one with the largest ratio of coefficients on real GDP to coefficient on credit-to-GDP.

Estimate the model with two lags by finding a posterior peak. Form impulse responses and error bands for them by posterior simulation. Discuss whether the two shocks are well determined by the data, and whether they have any plausible substantive interpretation.

Note that the time series are of different lengths across variables and countries. Truncate each country's data so all variables are well defined. this will produce different length time series for each country, but this should cause no problems in stacking up the data across countries. Notice that you will need to use the "breaks" parameter in the R function (or something similar in your own code) to avoid having "lags" pick up data from other countries in the stacked data.

Do what you can to assess whether the assumption that $A(L)$ is constant across countries is consistent with the data.

This work is meant to take up no more than one long day's effort. You are not to discuss it with other students while you are working on it, though you can request clarifications from me by email. If the work proves to be too much for one full day's effort, write up as much progress as you can achieve in that time. A full day's work of intelligent effort is likely to get an A even if incomplete.

Code and a data file are available at `sims.princeton.edu/yftp/ApplEmet14/TH`