BAYESIAN ANALYSIS OF THE EMPLOYMENT RATIO REGRESSION

We will consider the same (now corrected) data as in the last exercise. The wage data should now be logged, to make interpretation of the regression easier, and this has been done in the .RData version of the data. The errors from some data having been read in as factors in the .RData version have been corrected.

There is a set of R files in the `blm` directory for this problem set that facilitate Bayesian regression for time series. The main programs are `tsregPrior`, which constructs a symmetric-in-variables prior that can favor persistence, as we discussed in class, and `blm`, which uses real data, the dummy observations created by `tsregPrior`, and hyperparameters for the prior on residual variance, to produce Bayesian posteriors. It returns the coefficients, standard errors, variance-covariance matrix, degrees of freedom for posterior variance, and residual sum of squares. It also returns the log of the marginal data density — the posterior integrated over the parameters.

By running `blm` on models with different variable lists or lag lengths, one can produce log posterior odds and posterior distributions across models. By sampling from the posterior on the coefficients and equation variances, one can generate draws from the posterior distribution of any desired function of the parameters. In this exercise, you will construct "step response functions" and a distribution for them.

The R code is in the form R uses for a "package". You *may* be able to, if you have copied the `blm` directory on the course web site into a `blm` subdirectory of your own working directory, give the command `install("blm")` in R, and then `library("blm")`, and then find R help working for these functions. If not, though, the information in the help files is all there, in somewhat less readable form, in the headers to the `.R` files. Though I have tested this code, it is still new, so errors are quite possible. Report errors or strange behavior to me by email, promptly. Extra credit for finding errors.

If you want to work in Matlab or some other language, you can duplicate the workings of the R files. The algebra involved is fairly straightforward, but getting the details working might be a lot of effort. If you've not used R before, note that it can be downloaded for free from the internet. If there's demand for it, I can run an R tutorial Wednesday or Thursday.

(1) Estimate the model with 3 lags of `emprat` and current and four lagged values of the other variables on the right-hand side. Start with `smooth=.7`, `damp=1.1`, `erratio=4`, `dfp=1`, and `scalepv=1`. Try at least 3 other settings for these hyperparameters and calculate the posterior probabilities of the settings you have tried. For the remainder of the exercise, use the hyperparameters with highest posterior probability, unless the estimates from this model look ridiculous.
(2) Estimate the model with all lags of one of the three independent variables completely excluded. Repeat for each of the three independent variables. Calculate the posterior odds across the four models — all included, and each exclusion.

(3) Using the model with all variables included, make 1000 draws from the joint posterior distribution of the residual variance and parameters, saving the results. The residual variance has an inverse-gamma distribution and the parameters, conditional on the residual variance, are joint-normal. So you can draw from the inverse gamma (or from the gamma, and then invert), and then draw from the joint normal. Then for each of these draws, calculate the step-response to a permanent unit change in each right-hand-side variable, saving the result. In R, this can be done via two applications of `filter()`. If the coefficients on the lagged dependent variable are in `by` and those on current and lagged `x` are in `bx`, this will do it, for example:

```
stpr <- filter(c(0,0,0, rep(1,45)), bx, type="convolve")
stpr <- filter(stpr[-(1:3)], by, type="recursive")
```

But it's also pretty easy to program this directly. The three initial zeros are so that, with three lags, the response starts up at time zero. The first three values from the first line are all "NA". Then, to show your results, plot the step response at the posterior mean, and also the 5th, 16th, 86th, and 95th percentiles of the distribution of your 1000 draws. Form these percentiles point by point over the horizon of the responses.