EXERCISE ON BREAKS

Using the same data on USGDP(not gdp growth) as in the previous exercise 1, in this exercise you form a posterior distribution across break dates, for disturbance variance and coefficients jointly. With y_t the log of GDP for quarter t, the model is

$$y_t = \rho_{1,i(t)} y_{t-1} + \rho_{2,i(t)} y_{t-2} + \alpha_{i(t)} + \varepsilon_t \tag{1}$$

$$\varepsilon_t \mid \{y_{t-s}, \text{ all } s < t\} \sim N(0, \sigma^2_{i(t)}) \tag{2}$$

$$i(t) = \begin{cases} 1 & t < B \\ 2 & t \geq B \end{cases} . \tag{3}$$

There is a set of R and matlab functions, with `mgnldnsty` being the main one that calls the others, on the course website along with this exercise. There is also an R function, `breaks()` in the `Breaks.R` file, that basically does the exercise for you, except for the plotting and thinking.

(1) Before doing any complicated estimation, plot the growth rate of gdp (the change in its log, which is the second column of the data file) against time. Based on the plot, form your own, approximate beliefs about whether there was a break in variance and a rough 90% probability interval for the range of dates at which a break occurred, assuming there is indeed exactly one break. Write these down for comparison to what emerges from the more formal data analysis.

(2) Plot the density function for the break date, using the default settings for the arguments of `breaks()`. Note that because `breaks()` (and `mgnldnsty` if you work with that directly) return the log of the marginal data density, the actual density exponentiates these values. The log density values may well be too big or small to be exponentiated without overflow or underflow numerical errors, so subtract a constant from them before exponentiating to keep the maximum value at a reasonable level. This of course only scales the density, so does not affect inference. Compare the modal break date and the implied 90% interval to your initial subjective estimates.

(3) The `breaks()` function has `ic = matrix(c(10,10,10)` as one argument. The prior used in the program includes dummy observations expressing the belief that forecasting $y(t+1) = \bar{y}$ when the lagged values of y are all $\bar{y}$ is likely to give a small error. The error size is implicitly approximately that of the residual error in the regression model. The default for `mgnldnsty.R` is `ic=NULL`, in which case $\bar{y}$ is taken to be mean of the initial conditions. Redo the calculations and plots with `ic=NULL` and discuss why they differ from the base case as they do. They should show a high probability of breaks at the beginning and end of the sample.

[The matlab version of `mgnldnsty` has no `ic` argument and (I think) always uses initial conditions in forming the prior. If you work in matlab, you can either try to

Date: March 29, 2015.

add an `ic` argument, or treat the prior on break dates as putting probability zero on breaks in the first or last two years.]

(4) Show the posterior modal parameters, conditional on a break at the modal break date, together with their standard errors. These can be constructed from the `$var` component of the list returned by `mgnldnsty` or `breaks`. E.g., the standard errors are `with(w1\$var, sqrt(sum(u^2) * diag(xxi)))`. The autoregressive coefficients are in `By` and the constant in `Bx`. Based on these estimates and standard errors, what is your judgement about whether it is likely that a formal odds ratio calculation would favor constant coefficients (but breaks in variance) over a model with breaks in both coefficients and variances?

(5) Still conditioning on the peak probability break date, use the `var` statistics to assess whether using differenced data (as in the original problem set 1) would have given very different results.

(6) Apply `mgnldnsty` to the full sample, with no break, and compare its value to the average across B of the marginal densities with breaks, to obtain an odds ratio for one break vs. no breaks.