MID-TERM EXAM

The exam is an extension of the model-choice exercise to a direct comparison between the "choice of model" framework of Fernandez et al (FLS) with a Bayesian ridge regression approach. You will also be comparing to methods for model choice: cross validation vs Bayesian posterior odds.

Ridge regression is usually characterized as replacing the usual $(X'X)^{-1}X'y$ formula for the OLS estimator with $(X'X + \lambda I)^{-1}X'y$, where λ is a tuning parameter. If the X variables have different variances, the λI term is replaced by λD, where D is a diagonal matrix with the diagonal elements of $X'X$ on its diagonal. The same effect can be obtained by normalizing all the X variables to have the same variance and using the λI form.

Ridge regression can also be arrived at by using a particular conjugate prior in the standard normal linear regression model. It is then possible to calculate the posterior marginal data density (mdd) analytically for any given value of λ and therefore, with a prior on λ an overall integrated mdd by integrating over λ.

In the previous exercise you already calculated the mdd for the FLS setup, but this was done with a flat prior over models. For the model comparison between ridge and FLS, a proper prior is required for each. A proper flat prior over the 2^{41} models is just 2^{-41} on each. So the previous calculation of the log of the integrated posterior has to be reduced by $41\log(2) = 28.42$.

For the cross-validation model choice part of this exam exercise, you will simply pick a subset of 24 the countries in the FLS data set to use as a test sample, and fit both the ridge and the FLS models to the remaining 48. Then you will predict the growth rates in the test sample using the fitted models. For the ridge model this is straightforward. For the FLS model, in principle you shouldmake forecasts with all the models that have non-zero probability and weight the results together using the models' posterior probabilities. This may not be computationally feasible. If so, you can sort the models by their posterior probabilities and use a feasible number of the high-probability models to weight together.

The prior for the ridge regression has to be proper in the residual variance σ^2 and the coefficients β jointly, so it should be the usual conjugate normal-inverse-gamma form. It can be implemented by using dummy observations in which $y = 0$ and $x = \sqrt{\mathrm{Var}(X_i)\lambda} e_i$, where e_i is a unit vector with 1 in the i'th place and 0 elsewhere. The prior on σ^2 can be imposed by using a single dummy observation in which $y = .01$ (for example) and $x = 0$. This implies an inverse-gamma distribution with scale parameter .0001 and shape parameter .5 (using the terminology of Gelman et al in BDA3).

While the mdd for a regression model with dummy observation prior can be computed analytically, accounting for all the terms in the normalizing constant is picky work. I need a program that does this for my own work and have it half-written. I'll post it to the course web site this morning. It will be `SNLMmdd(Y,X)` and will return the log of the integral of

Date: March 13, 2020.

the SNLM likelihood for the given Y and X. To use it to evaluate the mdd of a posterior, when you have the first `ntrain` observations at the top of Y and X as dummy observations, you would get the log mdd for the full Y, X, then subtract the log mdd for Y, X truncated to include only the dummy observations.