Computer Tutorial 2: Bayesian Analysis of the Regression Model

Data and Matlab code for all questions are available on the course website. The first three of these exercises use the data set HPRICE.TXT which is explained in the textbook (Koop, 2003, pages 47-48). In this data file the first column is the dependent variable and the remaining columns are explanatory variables and the code uses the first four explanatory variables (in columns 2 through 5 of HPRICE.TXT). The fourth exercise uses the cross-country growth data set (growth.dat).

Exercise 1: *The Normal Linear Regression Model with Natural Conjugate prior (Analytical results)*
Bayesian inference in the Normal linear regression model with natural conjugate prior can be done analytically. That is, analytical formulae for the posterior exists. Construct a program which uses these analytical formulae to calculate the posterior mean and standard deviation of the regression coefficients for an informative and a noninformative prior. This is done in Session2_Ex1.m. Optional exercise: Extend this code to calculate marginal likelihoods and calculate posterior odds ratios for $\beta_j = 0$ for each regression coefficient.

Exercise 2: *The Normal Linear Regression Model with Natural Conjugate prior (Monte Carlo integration)*
Repeat Exercise 1 using Monte Carlo integration. How many draws to you need to take to replicate your answers to Exercise 1? How sensitive are results to your choice of the number of draws? Code is available in Session2_Ex2.m.

Exercise 3: *The Normal Linear Regression Model with Independent Normal-Gamma prior (Gibbs sampling)*
Bayesian inference in the Normal linear regression model with independent Normal-Gamma prior requires Gibbs sampling. Repeat Exercise 2, but using an independent Normal-Gamma prior and, thus, using Gibbs sampling. How do your results compare to Exercise 1 and 2? Code is available in Session2.Ex3.m.

Exercise 4: *BMA in Cross Country Growth Regressions*
In my handout, "Bayesian Methods for Fat Data", I carry out a BMA exercise using a data set from the economic growth literature. This data set covers $N = 72$ countries and contains $K = 41$ potential explanatory variables. The dependent variable is average per capita GDP growth for the period 1960-1992. In it, I use a g-prior value of $g = \frac{1}{K^2}$ recommended by Fernandez, Ley and Steel (2001, Journal of Econometrics) and the MC^3 algorithm took $2,200,000$ draws and discarded the first $200,000$ as burn-in replications. Repeat this example using smaller numbers of draws so as to iinvestigate how sensitive results are to the number of draws. Next investigate the sensitivity of results to the choice of g. Code is available in BMA.m. Optional exercise: For a given value of g, the weighted average of the marginal likelihoods across the draws can be used as a measure of support for the value of g. Modify the code to calculate this measure of support for a range of values of g, selects the value of g which maximizes it and presents BMA results for this value.