Computer Tutorial 2: Bayesian Analysis of the Regression Model

Data and Matlab code for all questions are available on the course website.

This computer tutorial takes you through a series of exercises which demonstrates Bayesian inference in the regression. The first three of the exercises go through the basics while the fourth puts it all together in an interesting time series application. If you are confident in your knowledge of the basics, then you might want to go directly to Exercise 4.

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: The AR(p) model as a Regression Model
This exercise is based on Geweke (1988, Journal of Business and Economic Statistics): "The Secular and Cyclical Behavior of Real GDP in 19 OECD Countries" which uses an AR(3) model for GDP growth:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \varepsilon_t$$

where ε_t is i.i.d. $N\left(0, h^{-1}\right)$. This can be treated a regression model (i.e. y_{t-1}, y_{t-2} and y_{t-3} play the role of explanatory variables).[1] Many important properties of y_t depend on the roots of the polynomial $1 - \sum_{i=1}^{p} \beta_i z^i$ which we will denote by r_i for $i = 1, .., p$. Geweke (1988) lets y_t be the log of real GDP and sets $p = 3$ and, for this choice, focusses on the features of interest: $C = \{\beta : \text{ Two of } r_i \text{ are complex}\}$ and $D = \{\beta : \min|r_i| < 1\}$ where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)'$. If the AR coefficients lie in the region defined by C then real GDP exhibits an oscillatory response to a shock and if they lie in D then y_t exhibits an explosive response to a shock. Note that C and D are regions whose bounds are complicated nonlinear functions of $\beta_1, \beta_2, \beta_3$ and hence analytical results are not available (even if a natural conjugate prior is used) and Monte Carlo integration or Gibbs sampling is required.
(a) Using an appropriate data set (e.g., the US real GDP data set provided on the website associated with this book), write a program which calculates the posterior means and standard deviations of β and $\min|r_i|$.
(b) Extend the program of part a) to calculate the probability that y_t is oscillatory (i.e., $\Pr(\beta \in C|y)$), the probability that y_t is explosive (i.e., $\Pr(\beta \in D|Data)$) and calculate these probabilities using your data set.

---

[1] To simplify things, it is common to ignore the (minor) complications relating to the treatment of initial conditions. Thus, assume the dependent variable is $y = (y_4, .., y_T)'$ and treat $y_1, .., y_3$ as fixed initial conditions.