Problem Set 6 – "Bayesian OLS"

In this problem set, use the Metropolis-Hastings algorithm to estimate the posterior distribution of the coefficients on returns to schooling.

In the repository you find data compiled by David Card that contains a sample with individual wage data. We are interested in finding the returns to schooling, ie the effect of an additional year in school on log wages. For the purpose of this problem set we ignore the endogeneity problems in such a regression! (But we could do everything using an IV strategy, it would only require us to use a different likelihood function.)

1. Run an OLS regression of log wages on education, experience, and dummy variables for SMSA status, race ("black"), and region ("South"). Find estimates for the coefficients $\hat{\beta}$ and their standard errors, as well as an estimate of the standard deviation of the residuals $\hat{\sigma}_\varepsilon$.

2. Now find the posterior distribution of the parameter vector $(\beta, \sigma_\varepsilon)$ by using the Metropolis-Hastings algorithm. For both of the following parts, you will have to experiment with the variance for the step size used in the creation of the new proposal value when creating the sample. A good idea is to set the standard deviation proportional to the standard errors of $\hat{\beta}$ and $\hat{\sigma}_\varepsilon$ you obtained above in the OLS regression. To get reasonably fast convergence properties, people sometimes shoot for an acceptance rate between 20% and 25%. Plot histograms of the posterior approximations you obtain.

   (a) Use a flat prior for all parameters.

   (b) You come across a study by someone else on an unrelated dataset that finds a value $\hat{\beta}_{educ} = 0.06$ with a confidence interval of $[0.035, 0.085]$. You think about using $\hat{\beta}_{educ}$ as the mean for a normal prior distribution for β_{educ}. For the standard deviation of your prior you decide to use2 times the standard error of $\hat{\beta}_{educ}$ in the previous study. Use a flat prior on all other parameters.

3. In a few sentences compare how the estimates from 1. relate to the posterior distributions in 2.a) and 2.b).