Economics 690
Problem Set: Intro to Gibbs Sampling

(1) This question will make use of the wage-education data set in the lecture notes and on the course website. Though we have not used it to this point, the data set also contains an ability (test score) variable. The ability variable is standardized so that positive values denote test scores above the mean, and negative values denote test scores below the mean.

From the raw ability data, create a binary variable D_i which equals one if ability is greater than or equal to zero (we will call this the "high" ability group), and otherwise equals zero.

We seek to investigate the possibility that log wage conditional *variability* differs across individuals with "high" and "low" test scores. To this end, we want to consider the model

$$y_i = \beta_0 + \beta_1 Ed_i + \beta_2 Ability_i + u_i, \quad u_i \overset{ind}{\sim} N(0, D_i\sigma_1^2 + (1 - D_i)\sigma_2^2).$$

Thus, σ_1^2 is the variance parameter for "high ability" people (with test scores at or above the mean) while σ_2^2 is the variance parameter for the "lower ability" individuals.

Use the Gibbs sampler to fit this model under

$$IG(3, 2.5)$$

priors for both σ_1^2 and σ_2^2 and also employ the prior

$$\beta \sim N(0, 4I_3).$$

Include a table listing all coefficient posterior means and standard deviations.

In addition, use the simulated output to provide a histogram of the parameter

$$\psi \equiv \frac{\sigma_2^2}{\sigma_1^2}.$$

What do you learn from this histogram? What is the probability that the "low ability" variance parameter exceeds the "high ability" variance parameter? (You also might want to think about how a Frequentist would address these types of questions, though you do not need to write anything down about this in your problem set).