ECNS 562 –Econometrics II
Eric Belasco
Homework 3
Due Monday, March 5

Be sure to include your R code, as well as typed out results.

Question 1. This question is concerned with simulating a censored data set and estimating it using OLS and Tobit methods. Consider the following model:

$$Y_i^* = X_0\beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \varepsilon_i, \text{ for } i = 1,\ldots,100$$

where X_0 is a vector of ones,
$X_1 \sim$ Uniform (0,1),
$X_2 \sim$ Uniform (0,3),
$[\beta_0, \beta_1, \beta_2] = [-3, 2, 0.5]$, and
$\varepsilon \sim$ Normal ($\mu = 0$, $\sigma = 3$)
$Y = \max(0, Y^*)$

(a) Simulate the given data (Y,X) based on the above parameters. Note that Y is the observable variable and Y* is latent.
(b) What is the degree of censoring? Plot separate histograms of Y and Y* using the hist() function in R. Discuss the fundamental differences between the variables as well as the histogram plots. Why are estimates of OLS estimates inconsistent?
(c) Estimate the simulated data using OLS and Tobit functions **that you coded**. Report and compare parameter estimates, t-ratios, and sigma squared estimates. Discuss the improvement in estimation in the Tobit model.
(d) Compute slope estimates in the Tobit model for the observable Y variable, rather than the latent variable (Y*).
(e) Adapt your Tobit function to perform estimation using a Hurdle model function. Conduct a likelihood ratio test on the data to assess which model is more appropriate. *(Hint: The Tobit model can be written as a restricted case of the Hurdle model)*
(f) Assume we wanted to account for heteroskedastic errors in Tobit our model and specified the following form: $\sigma_i^2 = \exp(X_i\alpha)$, where $\alpha = [\log(\sigma^2) \quad \alpha_1 \quad \alpha_2]$.
   i. Simulate a new set of data that incorporates this heteroskedastic structure with true values (of your choosing) for the three new parameters.
   ii. Write out the original likelihood function for the Tobit model and modify it to account for the above form.
   iii. How can parameter estimates for α_1 and α_2 be interpreted? If these variables are statistically significant, what does this infer about the data?
(g) Would the above simulated data be more accurately characterized by a two-part model? Discuss why or why not and the notable differences between the Tobit model and other models for censoring (Hurdle, Double Hurdle, etc.).