Econ 342 February 11, 2009
Shakeeb Khan Due: March 2, 2009
Spring 2009

Problem Set 4

1. (MATLAB question) In this question we will explore determinants of marriage length by using data from the June 1985 CPS (current population survey) marriage and fertility supplement. This data set was obtained from interviews with women aged over 18, and is comprised of the following variables :

   (a) First column = age at first marriage in months

   (b) Second column =age at termination of first marriage in months

   (c) Third column = age at time of interview in years

   (d) Fourth column = race: 1,2,3 corresponding to white, black, other, respectively

   (e) Fifth column = education in years

   (f) Sixth column = number of children

   (g) Seventh column = income bracket

   For this question, we will treat marriage length as the dependent variable. One immediate problem we face is that this dependent variable is *censored*. By this I mean that the interviewee is still married at the time of the interview, so we do not know how long their marriage will last. For these people, their age at termination is listed as 0 in the data set. For now, we will deal with problem in two ways (neither of which, it turns out, are totally satisfactory. We'll learn how to appropriately deal with censored data in the second semester or second year, so for now but we'll simply see if the regression coefficients we get make any sense.)

   (a) First, let's "throw away" the censored observations, and run the regression with race, education, number of children and income bracket as the explanatory variables. Assuming the normal regression model, test the significance (at the 5% level) of each of the regression coefficients, against two sided alternatives.

   (b) Second, we put "lower bounds" on the dependent variable for censored observations. That is we compute the length of marriage for censored observations as age at time of interview minus age of first marriage. (Make sure the units of

measurement match up). With this data set, do the same exercises you did in the previous question.

(c) Finally, conduct the same exercise now using Tobit MLE.

2. Hayashi Chapter 7, page 496, #1, #2.

3. Let $y_i = \max(x_i'\beta_0 + \epsilon_i, 0)$. Formally derive the expression for $E[y_i|x_i, y_i > 0]$ and $E[y_i|x_i]$ assuming ϵ_i is normally distributed with variance σ_0^2.

4. Now suppose we alter the basic censored regression model in the following way:

$$y_i^* = x_i\beta_0 + \epsilon_i \quad \epsilon_i \sim N(0, \sigma_0^2) \tag{1}$$

$$y_i = cI[y_i^* \leq 0] + y_i^* I[y_i^* > 0] \tag{2}$$

where c is some constant.

(a) Find $E[y_i|x_i, c]$

(b) How does the MLE change with c?