Economics 471: Applied Econometrics
Problem Set 5

This problem set will be due on the last day of class, May 2, and we will have our usual in-class presentations of it.

This problem concerns predicting productivity of new workers in a large American manufacturing firm. The data will be available on the class website with the name `weco.dat`. There are seven variables: y_i – an observed standardized physical productivity measure for the i^{th} worker after the initial training period, sex_i – a dummy variable for the workers' sex (males are 1) dex_i – a score on a physical dexterity exam administered before the worker was hired, lex_i – the number of years of education of the worker, and *quit* – whether the person quit within the first six months (quitters are 1). The last two columns of the data provide actual duration of employment and a censoring indicator, respectively. If the censoring indicator is 0 then the corresponding duration is censored. The duration variable is not needed for the analysis proposed in the questions below.

1. Estimate the model
$$y = \alpha_0 + \alpha_1 sex + \alpha_2 dex + \alpha_3 lex + \alpha_4 lex^2 + u.$$
   (a) Test the hypotheses: $H_0 : \alpha_3 = \alpha_4 = 0$ and $H_0 : \alpha_4 = 0$. Interpret the results of these tests in economic terms.
   (b) Given the results of part a) draw a diagram illustrating the dependence of "mean productivity" on education. For this purpose, set dexterity at its mean and $sex = 0$. Interpret the picture. How does it change for men? Suppose you thought the *shape* of the education effect was different for men and women; reestimate your respecified model. Does this change your conclusions?

2. Now consider the possibility that the *dispersion and perhaps even the shape* of the conditional density of productivity depends on the $sex - dex - lex$ variables. Consider a quantile regression model of this type, estimate and interpret it. For this purpose, redoing the prior plots of mean productivity for several quantiles would be helpful.

3. Now consider a similar model for quits
$$logit(P(quit = 1)) = (\beta_0 + \beta_1 sex + \beta_2 dex + \beta_3 lex + \beta_4 lex^2)$$
   where $quit = 1$ if the worker quit within the first 6 months after employment, and is 0 otherwise. Presumably, there is a fixed cost of hiring and training so there is an incentive on the part of the firm to avoid hiring workers who are likely to quit after only a few months. How would your findings be expected to influence the firms willingness to hire workers of various education levels?

(a) Estimate this model by logit, interpret the estimated parameters, in particular the estimated education effect. Draw a picture as in part (1b.) above of the probability of quitting as a function of years of education. Explain the connection between the parameter estimates and the picture.

(b) Explore the effect of gender along the lines of question 1b.

4. Finally, we wish to reconsider the *sexdexlex* productivity model of Question 1 exploring the consequences of "sample selectivity". Suppose instead of observing the entire sample of 683 individuals, we instead observed productivity only for those who didn't quit.

   (a) Use the Heckman two-step procedure to estimate the productivity equation of Question 1, using only this "censored" data.

   (b) Compare and contrast these results with your previous results using the full sample, and the results from (naively) applying OLS to the censored sample. In particular, discuss how the inferences drawn above are altered by the sample selection of non-quitters.