ECON 407: Problem Set 4 April 7, 2013

This problem set is due on the Friday of the last week of class (April 25, 5pm)). Instructions:

- Only hardcopies will be accepted.
- Late work will not be accepted under any circumstances.
- As usual, incorporate Stata commands, results, and narrative in your responses to this Problem Set. Combine these into a continuous narrative so I don't have to jump around looking at code and reading results.
- Celebrate that this is your last Cross-section problem set.

Consider the data from a recent study by Deb, Munkin and Trivedi (2006) on health expenditures. The paper can be found by clicking here. The methods they use in the paper are on the cutting edge of econometrics research, and are way beyond the scope of the course. We will be using their data for estimating Tobit and Heckman models we investigated in class. As difficult to read as the paper is, you may find their discussion useful. The data we use differs from that used in the paper by only including year 2001.

Access the data by issuing the commands
`webuse set http://rlhick.people.wm.edu/econ407/data` and then
`webuse deb_munkin_trivedi`.

Table 1: Health Expenditure Data.

Variable	Description
ambexp	Ambulatory medical expenditures (excluding dental and outpatient mental)
lambexp	=ln(ambexp) given ambexp > 0 ; missing otherwise
dambexp	=1 if ambexp > 0 and 0 otherwise
lnambx	=ln(ambexp) if ambexp>0 and 0 if ambexp=0
ins	=1 if enrolled in health insurance (PPO or HMO), 0 otherwise
totchr	Number of chronic diseases
age	Age in years / 10
female	=1 if female, otherwise 0
educ	Years of schooling
blhisp	=1 if black or hispanic, 0 otherwise

1. The Tobit Model:
   (a) Estimate an OLS model of medical expenditures. Take care to plot your dependent variable to examine potential truncation or censoring points. Provide a histogram of 'ambexp' and 'lambexp'. Which looks more like a normal distribution? Are there any practical downsides to the ln transformation?
   (b) Recast your model as a censored regression model (and suitable for Tobit) by recoding missing values of 'lambexp' according to some minimum value for ambexp slightly greater

than zero (since ambexp=0 for some observations, the lambexp is undefined). Note: see the Tobit/Heckman code I provided for an example of this. Provide a figure depicting where the censoring point is, and show the probability of *being selected* as an area under the pdf curve. How many observations are censored?

(c) Write down the equation for the expected value of health care expenditures in a Tobit framework. What is the role of the inverse of Mill's ratio in this expression?

(d) Given your coefficients in the previous part, calculate the inverse Mill's ratio for sample. How does this value vary according to actual health care expenditures and your censoring point for the first 10 or so observations?

(e) Interpret and describe the various marginal effects you get from the Tobit model. To fully answer this question, you might need to discuss the underlying expected value being used in the calculation (e.g. E(Y|Y>a), E(Y*), E(Y), Prob(Y>a)).

(f) Calculate the log-likelihood function evaluated at the model estimates.

2. Comment on the suitability of the OLS, the Tobit, and the Heckman model for this problem.