Econ 240B Spring 2009
Problem Set 4

This problem set is due in class on Monday April 27th, 2009

1. You can use any statistical software to answer the questions, including Stata, Sas, Eview and Tsp. Or you can use Gauss or Matlab.

   The purpose of this exercise is to introduce procedures for limited dependent variable models. The data set for this exercise is that subset of the NLSY used by Herrnstein and Murray in their book, The Bell Curve. The dataset and the instructions are available from the class page. Please copy the data to your own directory before working on it. Turn in the relevant computer output and written materials to answer the following questions. The documentation for the variables are in belldat.doc.

   Consider the following model:

$$\ln W = X_w \beta_w + U_w \tag{1}$$

   where W is the women's weekly wage rate and the vector X_w contains an intercept, education(E), age(A), SES, AFQT, race, A^2, $A * E$, and $AFQT * E$. The utility of work over staying at home is determined by the equation:

$$Utility = X_{wk}\gamma_{wk} + \gamma_w \ln W + V = X\beta_{wk} + U_{wk} \tag{2}$$

   where X_{wk} contains an intercept, family income, and a dummy variable indicating whether the woman is currently married or not. The vector X contains all of the unique elements of X_w and X_{wk}. The discrete variable $\delta = 1$ when $Utility > 0$ and $= 0$ when $Utility \leq 0$.

   Use the data set bell_female.dat.

   (a) Estimate the parameters determining labor force participation assuming that the error $U_{wk} \sim N(0, 1)$ in model (2). Test the hypothesis that marriage influences participation using both Wald and likelihood ratio test statistics.

   (b) Instead of using the probit model, now estimate the parameters determining labor force participation assuming that the error U_{wk} follows a logistic distribution. Test the hypothesis that marriage influences participation using both Wald and likelihood ratio test statistics. Are the results from the logit and probit models very different?

   (c) Consider a three-state classification of a woman's hours of work: she doesn't work at all(designate by setting the discrete variable b=1); she works part of the year which implies that $0 <$ Weeks < 20(designated by b=2); and she works most of the year with Weeks ≥ 20(b=3). Formulate a multinomial logit model and estimate the probability $Pr(b = j|x)$ using this multinomial logit model. This is a different model than (2).

(d) What is the change in the probability of working part of the year when marital status changes from not married to married? Is this change statistically significant? (You need to calculate the standard error of the predicted change.) First answer this question by holding all the other variables in X at their means. Then answer this question again by holding all the other variables in X at their medians.

(e) Assume that the errors (U_w, U_{wk}) follow a bivariate normal distribution. Data are available on wages only for workers. Estimate the coefficients β_w of model (1) accounting for sample selection using MLE

2. Consider the model:

$$y_1(t) = x_1(t)\beta_1 + g_t e_1(t)$$
$$y_2(t) = x_2(t)\beta_2 + e_2(t)$$

where

$$\begin{pmatrix} e_1(t) \\ e_2(t) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & 1 \end{bmatrix}\right)$$

and $g_t \equiv g(x(t), \gamma) > 0$ is a known function depending on an unknown parameter γ. Let:

$$\delta(t) = \begin{cases} 1 & \text{when} \quad y_2(t) > 0 \\ 2 & \text{when} \quad y_2(t) \leq 0 \end{cases}$$

(a) Suppose you are given only data on $\delta(t)$. Describe the estimation of β_2 using maximum likelihood. Explicitly write out the objective function you would use and list the properties of the estimator.

(b) Suppose you want to estimate β_2 using NLLS applied to the model:

$$\delta(t) = E(\delta(t) | x(t)) + v(t).$$

Explicitly write out the objective function used to carry out this estimation. What are the asymptotic properties of the NLLS estimator? Would you use the robust standard error option, and if so why? Other than maximum likelihood estimator, is there a more efficient estimator than the NLLS estimator? If so describe it in detail.

(c) Suppose $y_1(t)$ is always observed. How would you estimate the parameters β_1 and γ using maximum likelihood methods? Explicitly write out the objective function used to carry out this estimation. What are the asymptotic properties of the estimators? Can one use comparisons of values of the likelihood function to test hypotheses involving the elements of β_1 without relying on the normality assumption? What about testing hypotheses involving restrictions on γ?