Professor R.L. Oaxaca	Assignment #2
Econ 522b	Fall 2009

Due Thursday, October 8

This assignment is on multiple outcome and limited dependent variable models. The necessary data are contained in the Excel files dat205a.xls and dat201b.xls and the STATA files dat205a.dta and dat201b.dta available from the website http://u.arizona.edu/~rlo. Be sure to attach the supporting computer print out to the completed assignment and make clear where your answers are shown.

The data for question 1 are provided in the file dat205a.xls or dat205a.dta.

1. Individuals traveling between two cities may choose among 4 different modes of travel: Air (j=1), Train (j=2), Bus (j=3), and Car (j=4). Individuals care about only the duration of the trip and the monetary cost of the trip. The utility function of a representative traveler can be expressed as

$$U_{ij} = \beta_1 Time_{ij} + \beta_2 Cost_{ij} + \varepsilon_{ij}$$

where $Time_{ij}$ represents the number of minutes it would take the ith individual to reach the destination using transportation mode j, $Cost_{ij}$ represents the monetary cost to the ith individual to use the jth transportation mode, and ε_{ij} is an i.i.d. error term from the extreme value distribution. Let Y_{ij} represent a dummy variable that takes on the value 1 if the ith individual chooses transportation mode j, and 0 otherwise.

a. Specify the likelihood function for the model and estimate the $\beta's$ by MLE.

b. Consider an individual who faces the travel time and travel cost situation described below.

Mode	Time	Cost
Air	120	90
Train	650	150
Bus	600	50
Car	550	60

(1) Estimate the probabilities that this individual would travel (a) by Air, (b) by Car.

(2) Estimate the elasticity of the probability of traveling by Car with respect to the cost of traveling by Car.

(3) Estimate the elasticity of the probability of traveling by Car with respect to the cost of traveling by Air.

The data for questions 2-5 are provided in the file dat201b.xls or dat201b.dta. The data consist of a national sample of 4,569 adults between the ages of 33 and 41 in 1998 and are drawn from the NLSY79 data set. For the exercises below, restrict the sample to the set of individuals for whom $grade98 > 0$, $female = 0$, and $other = 0$.In this data set $grade98$ is the highest grade completed, $female$ is an indicator variable for gender, $other$ is an indicator for race other than white or black, $black$ is a race indicator variable for black, and $union98$ is a indicator variable for union membership.

2. Consider the following model of job satisfaction

$$\begin{aligned} Y_i^* &= \beta_0 + \beta_1 black_i + \beta_2 union98_i + \sum_{j=1}^{4} \beta_{j+2} oc_{ji} + \varepsilon_i, \\ i &= 1, ..., 2207 \end{aligned}$$

where Y_i^* is a latent variable, ε_i is an i.i.d. logistically distributed random variable, and oc_j is an occupational indicator variable defined by

$$\begin{aligned} oc_4 &= 1 \text{ if } occup = 4 \text{ (professional/managerial)} \\ &= 0 \text{ otherwise} \\ oc_3 &= 1 \text{ if } occup = 3 \text{ (sales/clerical)} \\ &= 0 \text{ otherwise} \\ oc_2 &= 1 \text{ if } occup = 2 \text{ (craft)} \\ &= 0 \text{ otherwise} \\ oc_1 &= 1 \text{ if } occup = 1 \text{ (service)} \\ &= 0 \text{ otherwise} \end{aligned}$$

Although Y_i^* is not observed, $jobsat_i$ is observed:

$$\begin{aligned} jobsat_i &= 0 \text{ if dislikes job} \\ &= 1 \text{ if likes job fairly well} \\ &= 2 \text{ if likes job very much} \end{aligned}$$

According to the model

$$\begin{aligned} jobsat_i &= 0 \text{ if } Y_i^* \leq 0 \\ &= 1 \text{ if } 0 < Y_i^* \leq \mu_1 \\ &= 2 \text{ if } \mu_1 < Y_i^*. \end{aligned}$$

a. Specify the likelihood function for this model and estimate its parameters by MLE.

b. Estimate the probability that a union white male professional worker would report that he likes his job very much.

3. Consider the following model of occupational employment:

$$\Pr\text{ob}(occ_i = 0) = \frac{1}{1+\sum_{m=1}^{2}\exp\left(X_i'\beta_m\right)}, \quad i = 1, ..., 2207$$
$$\Pr\text{ob}(occ_i = j) = \frac{\exp\left(X_i'\beta_j\right)}{1+\sum_{m=1}^{2}\exp\left(X_i'\beta_m\right)}, \quad j = 1, 2$$

where $j = 0, 1, 2$ correspond to (1) laborers, operatives, and service workers; (2) craft workers, sales, and clerical workers; and (3) professional and managerial workers. The vector $X_i' = (1, black_i, union98, grade98_i)$ The variable *occ* is defined as follows:

$$\begin{aligned} occ &= 2 \quad \text{if } occup = 4 \\ &= 1 \quad \text{if } occup = 2 \text{ or } 3 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

a. Specify the likelihood function for the model and estimate the model's parameters by MLE.

b. Estimate the marginal effects of the highest grade completed on the probabilities of employment in each occupational group for a union, black male with 12 years of schooling.

The following information pertains to questions 4 and 5.

Consider the true model: $autoval_i = \beta_0 + \beta_1 black_i + \beta_2 \ell n(ntinc_i) + \varepsilon_i$, $i = 1, ..., 2207$, where *autoval* is the market value ($1,000's) of all vehicles currently owned by the individual, *ntinc* is the individual's family net income ($1,000's), $\varepsilon_i \sim N(0, \sigma^2)$ and satisfies all of the classical assumptions.

4. Suppose one had access only to the observations for which $autoval_i > 0$.

a. Use MLE to estimate the model

$$(autoval_i \mid autoval_i > 0) = \beta_0 + \beta_1 black_i + \beta_2 \ell n(ntinc_i) + (\varepsilon_i \mid autoval_i > 0).$$

b. Estimate the marginal effect on $E(autoval \mid X', autoval > 0)$ of net family income for a white male automobile owner with the mean family net income of white male automobile owners.

5. Suppose one observes Y_i rather than $autoval$, where

$$\begin{aligned} Y_i &= 0 & \text{if } & autoval_i \leq 0 \\ &= autoval_i & \text{if } & autoval > 0 \end{aligned}$$

   a. Use a MLE method to estimate the $\beta' s$
   b. Estimate the marginal effect on $E(Y \mid X')$ of net family income for a white male with the mean family net income of the total white male sample.

6. A researcher wishes to estimate a simple model of wage determination in the union sector:

$$ln(earnings_i) = \beta_0 + \beta_1 black_i + \beta_2 grade98_i + \varepsilon_i, \;\; \{i|\; union98_i = 1\}$$

where *earnings* is the individual's annual earnings ($1000's). The researcher is concerned that there may be selection effects in terms of the kinds of workers who are unionized. Accordingly the researcher specifies a union selection equation:

Prob($union98_i = 1$) = Prob($u_i > -\mathrm{I}_i$)

where $\mathrm{I}_i = \gamma_0 + \sum_{j=1}^{4} \gamma_j oc_{ji} + u_i$.

Assume that ε_i and u_i satisfy all of the classical assumptions and follow a bivariate normal distribution $(0, 0, 1, \sigma_u^2, \rho)$.

   a. Use the Heckit 2 step method to obtain consistent estimators of the parameters γ and β.
   b. Use a consistent estimator to estimate var($\varepsilon_i \mid union98_i = 1$) using the residuals from the estimated union earnings equation that controls for the union selection process.
   c. What do the empirical results suggest about the unobserved determinants of earnings of unionized workers? Explain.
   d. Estimate the incremental log earnings effect for a unionized worker of being a professional worker compared with a clerical worker.