Labor Economics
Assignment 3
Fall 2000

Please complete as much of this assignment as you can by Wednesday, November 1. Feel free to e-mail me with questions over the next week. I forgot to make copies of the Wales and Woodland paper that is on your reading list, but hopefully will be able to get copies made and placed in my 3rd floor mailbox by Tuesday, October 24. If they are not there please send me an e-mail!

You are to investigate the sensitivity of inferences to the estimator used when estimating labor supply functions in which there is a strong *a priori* suspicion of nonrandom sampling. The data you are to use are contained in the GAUSS data file `lsup_1.dat`, and the description of the data follows in the attached codebook.

The majority of sample members, married women between the ages of 25 and 34 in the March 1998 Current Population Survey, are not working at the time of the interview. You are to implement 3 different estimators, all based on the following model specification:

$$h_i^* = \alpha_1 + \alpha_2 \ln w_i + \alpha_3 \text{age}_i + \varepsilon_{1i} \quad (0.1)$$

$$\ln w_i = \beta_1 + \beta_2 \text{age}_i + \beta_3 \text{hs\_dip}_i + \beta_4 \text{college}_i + \varepsilon_{2i}, \quad (0.2)$$

where the α and β vectors are unknown parameters, ε_i has a bivariate normal distribution with mean vector $(0\ 0)'$ and covariance matrix Σ_ε, and h_i^* denotes "desired hours," which are the solution to the utility maximization problem without imposing the constraint that labor supply be nonnegative (i.e., that doesn't impose the constraint that leisure consumption be no more than the time endowment T).

1. Estimate equation (0.1) using information only from the subsample of working women and an OLS estimator.

2. Define the dependent variable $d_i = 1$ if $h_i^* > 0$ and $d_i = 0$ if $h_i^* \leq 0$. Define the maximum likelihood estimator of the employment decision (i.e., d_i), and implement it using the maxlik procedure in GAUSS to obtain estimates of all identified parameters. [HINT: This will be a probit estimator with only certain functions of the vectors α and β identified. Begin by substituting [0.2] into [0.1] and recognize that the disturbance in the new "desired hours" equation is normally distributed.]

3. Use your estimates from question 2 to implement the Heckman two-step estimator. In question 2 you estimated the function $Z_i(\delta/\sigma_v)$, where Z_i was the vector of exogenous variables in (0.1) and (0.2), δ is the vector of corresponding parameters from α and β, and σ_v is the standard deviation of $(\varepsilon_{1i} + \alpha_2\varepsilon_{2i})$. Using your estimates of this function, form $q_i = \phi(Z_i(\widehat{\delta/\sigma_v}))/1 - \Phi(Z_i(\widehat{\delta/\sigma_v}))$. Using only the sample of working women, use OLS to estimate

$$h_i = \alpha_1 + \alpha_2 \ln w_i + \alpha_3 \text{age}_i + \alpha_4 q_i + u_i,$$

where u_i is a disturbance term that is mean-independent of the right hand side variables.

4. Assume that the covariance between ε_{1i} and ε_{2i} is 0. Form the conditional likelihood estimator of (0.1) using only the subsample of working women. The conditional log likelihood is given by

$$\begin{aligned} L_c \quad = \quad & \sum_{\{h_i>0\}} \ln\{\frac{1}{\sigma_{\varepsilon_1}}\phi(\frac{h_i - \alpha_1 - \alpha_2 \ln w_i - \alpha_3 \text{age}_i}{\sigma_{\varepsilon_1}})/ \\ & (1 - \Phi(\frac{h_i - \alpha_1 - \alpha_2 \ln w_i - \alpha_3 \text{age}_i}{\sigma_{\varepsilon_1}}))\}, \end{aligned}$$

where ϕ is the standard normal p.d.f. and Φ is the standard normal c.d.f. In estimating this model, be sure to parameterize $\sigma_{\varepsilon_1} = \exp(\alpha_4)$, say, so that the maxlik procedure doesn't try to search for values of the standard deviation parameter that are negative.

5. Compare the point estimates of α_2 that you were able to obtain using the various estimators. Which seems more believable, if any? What are features of the data are being exploited by the various estimators that could account for the differences in the point estimates?

6. Write down the full information maximum likelihood estimator (FIML) for the two equation model [HINT: It has a "Tobit" structure.] Are all of the parameters of the model identified using this estimator, in particular, all of the elements of Σ_ε? [If you are very ambitious and/or have the time, try implementing it.]

CODEBOOK
DATA SET: lsup_1

Column	Variable	Values
1	*hours*	≥ 0
2	*lnwage*	ln wage rate if $h > 0$ 0 if $h = 0$
3	*age*	25-34
4	*hs_dip*	1 if completed exactly 12 years of schooling 0 if not
5	*college*	1 if completed more than 12 years of schooling 0 if not