

Econ 673 – Spring 2008  
Final Examination

1. Critically evaluate each of the following claims, explaining why each statement is either true or false. (10 points each)

- a) You are informed that two countries (A and B) are similar in terms of household bicycle ownership. Specifically, the number of bicycles ( $y_{ij}$ ) owned by household  $i$  in country  $j$  is known to follow a Poisson distribution, with

$$E(y_{ij} | X_{1ij}, X_{2ij}) = \exp(\beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij}); i = 1, \dots, N_j, j = A, B, \quad (1)$$

where  $X_{1ij}$  denotes the number of individuals in household  $i$  and  $X_{2ij}$  denotes the household's income. Note that the coefficients are assumed to be the same in the two countries. It is also known that the average number of bicycles per household is higher in country A than in country B. From this information, you can conclude that the marginal impact of household size on the average number of bicycles owned per household is larger in country A than in country B.

- b) Given the strong ignorability assumption, the average treatment effect (ATE) is the same as the treatment on the treated (TT).

2. (25 points) You have been asked to study the labor market participation of couples in Iowa. Specifically, you are told that the labor market participation decision of the male in household  $i$  is determined by the latent variable:

$$y_{Mi}^* = \alpha_M + \beta_M' X_{Mi} - \varepsilon_{Mi}. \quad (2)$$

Similarly, the labor participation decision of the female is determined by the latent variable

$$y_{Fi}^* = \alpha_F + \beta_F' X_{Fi} - \varepsilon_{Fi}. \quad (3)$$

Much to your relief, you are informed that it is well known that in Iowa  $\varepsilon_{Mi} \sim N(0,1)$  and  $\varepsilon_{Fi} \sim N(0,1)$  and, moreover, that the two error terms are independent of each other. You have available to you a random sample of  $N$  households and all of the necessary explanatory variables (i.e.,  $X_{Mi}$  and  $X_{Fi}$ ). Unfortunately, you discover that available data does not indicate which household member is working, but only whether none, both, or only one member in the household is working. That is, you only know  $y_i = y_{iM} + y_{iF}$ , where

$$y_{ij} = \begin{cases} 1 & y_i^* \geq 0 \\ 0 & y_i^* < 0 \end{cases} \quad (4)$$

is the dummy variable indicating that individual  $j (=F, M)$  in household  $i$  is working. Specify the log-likelihood function that you would need to employ to obtain maximum likelihood estimates of the parameters of your model.

3. (25 points) Suppose that you have data on the number of students enrolled in the undergraduate econometrics courses taught during 2007 at all of the major universities in the U.S. You are asked to analyze undergraduate interest in econometrics as a function of
- the timing of the course (e.g., say using  $X_{it} = |h_i - 12|$  where  $h_i$  denotes the hour of the day;
  - the textbook used (say using dummy variables for the various texts); and
  - the experience of the instructor (e.g., in years since PhD).

It is suggested that a count data model (and particularly a Poisson regression model) would be best suited to this problem. Unfortunately, you also realize that at each university course enrollment was limited to no more than 20 students and no fewer than 5 students in each semester.

- a) Specify the appropriate log-likelihood function that you would use to estimate the Poisson regression model's parameters.

- b) As an afterthought, you recall from your Econ 673 course that overdispersion can be a potential problem in count data model applications. Your classmate, however, suggests that you need not worry about this problem in the current case, since the unconditional variance of the enrollment figures is only slightly (10%) larger than the unconditional mean of the enrollment figures. Does this information reduce your concern regarding overdispersion? Explain why or why not.

4. (30 points) Consider the demand for airline travel, determined by the latent demand equation:

$$M_i^* = \alpha + \beta A_i + \delta Y_i + \varepsilon_i \quad (5)$$

where  $A_i$  denotes the age of the individual and  $Y_i$  denotes the individual's income, with  $\varepsilon_i \sim N(0, \sigma^2)$ . Suppose you are given access to data on a random sample of frequent flyers (i.e., those traveling more than 100,000 miles per year), including the number of miles flown, their age, and their income level.

- a) Can you use a two-step Heckman procedure, along with this database, to estimate the demand relationship in (5)? If so, how would you implement it? Be specific regarding the steps involved. If not, why not?

- b) Can you use maximum likelihood procedures with this database to estimate the demand relationship in (5)? If so, how would you implement it? Be specific regarding the log-likelihood function you would use.

- c) Finally, the airline industry reveals that it has a second database, consisting of a random sample of the general population indicating whether or not they fly; i.e.,

$$M_i = \begin{cases} 1 & M_i^* > 0 \\ 0 & M_i^* \leq 0. \end{cases} \quad (6)$$

The new dataset also contains the corresponding age and income information for each individual. Can you use the new data (together with the original data set) to estimate the model in (5) via the two-step Heckman approach?