Econometrics, Semester II, 2024-25
Homework I (100 points)

Instructor: M.A. Rahman

Deadline: 5:00 pm, February 5, 2025.

Please read the instructions carefully and follow them while writing answers.

- *Solutions to homework should be typed in LaTeXor written in A4 size loose sheets.*
- *Questions should be answered in order as they appear in the homework. Every new question should begin in a new page. Please number all the pages of your homework solution.*
- *Please leave a margin of one inch from top and one inch from left. Staple the sheets on the top-left.*
- *Matlab assignments (if any) and written answers should be together and in order.*
- *Please write your name and names of your group members on the first page of your answer script.*

---

1. (5+5+15 = 25 points.). Consider the binary logit model expressed in terms of a continuous latent random variable z_i as,

$$\begin{aligned} z_i &= x_i'\beta + \epsilon_i, \qquad\qquad \forall\, i = 1, \cdots, n, \\ y_i &= \begin{cases} 1 & \text{if } z_i > 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{1}$$

where x_i is a $k \times 1$ vector of covariates, β is a $k \times 1$ vector of unknown parameters, n denotes the number of observations, and the error follows a standard logistic distribution $\epsilon_i \sim L(0, \pi^2/3)$ for $i = 1(1)n$. Based on the logit model, answer the following.

(a) Find the probability of success $\Pr(y_i = 1)$?

(b) Derive the likelihood function of the logit model.

(c) Consider a study in which the dependent variable is the probability that the subject dies before age 65, and the primary explanatory variable of interest is whether the person smoked (at all)

in the years prior to age 65. Let Smoke_{i2} be an indicator for smoking status, and β_{smoke} be the corresponding coefficient. Then the latent regression in equation (1) becomes:

$$z_i = \beta_1 + \text{Smoke}_{i2}\beta_{\text{smoke}} + x_{i3}\beta_3 + \ldots + x_{ik}\beta_k + \epsilon_i$$

Find the odds of mortality by age 65 if individual i was a smoker ($\text{Smoke}_{i2} = 1$) and the odds if individual i was a nonsmoker ($\text{Smoke}_{i2} = 0$). What is the log-odds ratio of mortality for a smoker vs nonsmoker?

2. (10+25+10 = 40 points). Consider the data in the file `TransportChoiceDataset.xlsx`. The objective is to study individuals choice between automobile and transit for trip to work. The dependent variable `depend` takes the value 1 if automobile is chosen and 0 if transit is chosen. The covariates in the model are `intercept`, `dcost`, `cars`, `dovtt` and `divtt`. A description of these variables is present in the file.

(a) Present the descriptive summary of the variables (i.e., mean and standard deviation for continuous variables and count and percentage for discrete variables) in a table.

(b) Estimate Probit and Logit models by regressing the dependent variable `depend` on `intercept`, `dcost`, `cars`, `dovtt` and `divtt`. Present the regression coefficients and the standard errors in a table. Numbers should be reported to 3 digits after the decimal. Interpret the coefficient for `cars`.

(c) Calculate the sum of the log-likelihood, Akaike Information Criterion, Bayesian Information Criterion and Hit-rate for the Probit and Logit models.

3. (5+5+5+10+10 = 35 points). Consider the data in the file "`Mroz Data.xlsx`", first utilized by Mroz (1987) to explore important issues of endogeneity and sample selection, but these will be neglected here to provide a simple example of Tobit model. The data corresponds to 753 married women for the year 1975 from the University of Michigan Panel Study of Income Dynamics. There are many variables in the file, but we will only employ a subset of the variables for our exercise.

Our response variable is the number of hours worked (`WHRS`). If we wish to estimate a model explaining the market hours worked by a married woman, what explanatory variables would we include? Factors that would tend to pull a woman into the labor force are her education (`WomenEduc`) and her prior labor market experience (`WomenExp`). Factors that may reduce her incentive to work are her age (`WomanAge`) and the presence of young children in the home i.e., number of children less than six years old at home (`childl6`).

Thus, we may propose the following regression model:

$$\texttt{WHRS}_i = \beta_1 + \texttt{WomenEduc}_i\beta_2 + \texttt{WomenExp}_i\beta_3 + \texttt{WomenAge}_i\beta_4 + \texttt{childl6}_i\beta_5 + \epsilon_i.$$

(a) Present a descriptive summary (mean and standard deviation) of the variables of interest. Report all results to two digits after the decimal.

(b) Estimate a linear regression model only on positive values of `WHRS` and report the coefficient estimates, standard errors, and t-values in a table. Are there reasons to believe that a linear regression framework will not be appropriate for this data? Please explain.

(c) Write down a Tobit model and the corresponding likelihood.

(d) Fit a Tobit model and report coefficient estimates, standard errors, and t-values in a table. Comment on the effect of each variable on the response variable.

(e) What is the marginal effect of on observed hours of work for another year of education? Assume `WomenEduc`, `WomenExp`, and `WomenAge` are set at the corresponding mean values and `child16` = 1.

References

Mroz, T. M. (1987), "The Sensitivity of an Empirical Model of Married Women's Hour of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765–799.