Bayesian Econometrics, ECO545A, Semester I, 2020-21
Homework II (130 points)

Instructor: M.A. Rahman

Deadline: 4:00 pm, November 23, 2020

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

- *Solutions to homework should be written in A4 size loose sheets. If you are not comfortable writing on white sheets, please ask for biology paper in Tarun Book Store.*
- *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 all sides. Staple the sheets on the top-left.*
- *Computational assignments (Matlab):* ***Please answer the questions and report the results. Codes should be attached as an appendix.***

---

1. **(5+10+15+5 = 35 points)** *Censored regression model (CRM) or Tobit Type I model*: Consider the file "`adoption.xlsx`", where the sheet "Data" contains the data and the sheet "Description" contains the description of the variables. The dependent variable is `winbet`. Transform the variable by subtracting \$125 so that the threshold for the CRM is 0. The independent variables are `Intercept`, gender of the horse (`male`, `female`, the base category is "gelding" i.e., "neutered"), color of the horse (`bay`, `black`, `brown`, `buckskin`, `dun`, `gray`, `pinto`, `redroan`, `sorrel`, the base category is a mix, called "other"), `age`, and `training`. Based on this data, answer the following.

   (a) Present a descriptive summary (mean and standard deviations) of the data in a table (No points will be given if it is not in a table).

   (b) Estimate a classical linear regression model, where the response variable is the modified winning bet (i.e., `winbet-125`) and the covariates are a constant and the remaining variables in the data set. Present the OLS estimates of β and σ^2.

   (c) Estimate a Bayesian CRM and report the posterior estimates (mean and standard deviations) obtained from 20,000 MCMC iterations after a burn-in of 10,000 iterations. To obtain the Bayesian posterior estimates, use the following prior distributions:

$\beta \sim N(0_k, 100 * I_k)$ and $\sigma^2 \sim IG(1/2, 1/2)$, where k denotes the number of covariates including the constant. Present the results in a table.

(d) What are the notable differences between the estimates from OLS and Bayesian CRM? Please explain in detail.

2. **(3 × 10 = 30 points)** In a survey conducted between March 13th to March 17th, 2013 by the Pew Research Center for the People & the Press by Abt SRBI (Schulman, Ronca & Buculvas, Inc.), individuals were interviewed over the phone (landline and cell phone) about their opinion on 'Marijuana Legalization' in the United States. Overall, a sample of 1,501 adult individuals were interviewed and their responses recorded. However, after cleaning the data, we have a sample of 1240 observation that is attached in the file 'marijuana.xlsx'. Based on the data, do the following.

Lets prepare the data for fitting a probit model using Classical and Bayesian methods.

(a) Recode the response variable $q85$ as $y_i = 1$ if 'Yes, legal' and $y_i = 0$ if 'No, illegal'.

Present the counts and percentages for each categories of y.

(b) Label the variable 'age' as x_2 and 'hh1' (number of members in the household) as x_3.

Present the mean, median, standard deviation, maximum, and minimum of 'age' and 'hh1'.

(c) Recode the variable 'past use' as 1 if 'Yes' and 0 if 'No'. Label the recoded variable as x_4. The variable 'past use' tells if the individual has used marijuana in the past.

Present the counts and percentages for each categories of 'past use'.

(d) Recode the variable 'sex' as $sex_i = 1$ if male and $sex_i = 0$ if female. Label the recoded sex variable as x_5. Similarly recode the variable 'parent' of children under the age of 18 as $parent_i = 1$ if 'Yes' and $parent_i = 0$ if 'No'. Label the recoded parent variable as x_6.

Present the counts and percentages for each categories of 'sex' and 'parent'.

(e) Lets work on marital status. Generate three categories as follows. Category 1 ('single') consists of 'Never been married', Category 2 ('uncouple') consists of 'Divorced', ''Separated' or 'Widowed' and Category 3 ('couple') consists of 'Married' or 'Living with a partner'. Create a variable x_7 which takes the value 1 if an individual belongs to the category 'single' and 0 otherwise. Similarly, create another variable x_8 which takes the value 1 if an individual belongs to the category 'uncouple' and 0 otherwise. We will use the category 'couple' as the base category and not include in the regression.

Present the counts and percentages of 'single', 'uncouple' and 'couple'.

(f) Lets work on the column 'income'. Create the category 'poorMiddle' by merging 'Less than 10000', '10 to under 20000', '20 to under 30000', '30 to under 40000' and '40 to under 50000'. Similarly, create the category 'comfortable' by merging '50 to under 75000' and '75 to under 100000'. Finally, create the category 'rich' by merging '100 to under 150000' and '150000 or more'. Now, create a variable x_9 which takes the value 1 if the individual belong the category 'poorMiddle' and 0 otherwise. Similarly, create another variable x_{10} which takes the value 1 if an individual belongs to the category 'comfortable' and 0 otherwise. We will use the category 'rich' as the base category and not include in the regression.

Present the counts and percentages of 'poorMiddle', 'comfortable' and 'rich'.

(g) Lets work on the column 'educ'. Create the category 'HSandBelow' by merging 'Less than HS', 'HS Incomplete' and 'HS'. Similarly, create the category 'lessThanBachelors' by merging 'Some college' and 'Associate Degree'. Finally create the category 'BachelorsandAbove' by merging 'Bachelors', 'Postgraduate Degree' and 'Some Postgraduate'. Now, create a variable x_{11} which takes the value 1 if the individual belong the category 'HSandBelow' and 0 otherwise. Similarly, create another variable x_{12} which takes the value 1 if an individual belongs to the category 'lessThanBachelors' and 0 otherwise. We will use the category 'BachelorsandAbove' as the base category and not include in the regression.

Present the counts and percentages of 'HSandBelow', 'lessThanBachelors' and 'BachelorsandAbove'.

(h) Lets work on the column 'race'. Create the category 'white' for white individuals, category 'black' for black individuals and category 'allOther' by merging the remaining races. Now, create a variable x_{13} which takes the value 1 if the individual belong the category 'white' and 0 otherwise. Similarly, create another variable x_{14} which takes the value 1 if an individual belongs to the category 'black' and 0 otherwise. We will use the category 'allOther' as the base category and not include in the regression.

Present the counts and percentages of 'HSandBelow', 'lessThanBachelors' and 'BachelorsandAbove'.

(i) Lets work on the column 'party'. Create the category 'democrat' for individuals whose party affiliation is 'Democrat', category 'republican' for individuals whose party affiliation is 'Republican' and category 'independentOthers' by merging the remaining party affiliations. Now, create a variable x_{15} which takes the value 1 if the individual belongs the category 'democrat' and 0 otherwise. Similarly, create another variable x_{16} which takes the value 1 if an individual belongs to the category 'republican' and 0 otherwise. We will use the category 'independentOthers' as the base category and not include in the regression.

Present the counts and percentages of 'democrat', 'republican' and 'independentOthers'.

Enough of text mining, lets get to real work!

3. (**5+5+5+5+10+10 = 40 points**) Consider a binary **probit model** and do the following.

(a) Derive the probability of $y_i = 1$ and $y_i = 0$. Construct the likelihood function.

(b) Estimate the binary probit model, with y as the response variable and $(1, x_2, \cdots x_{16})$ as the covariates.

(c) Calculate the covariate effect of increasing the age by 5 years on $\Pr(y_i = 1|\beta)$.

(d) Calculate the covariate effect of being a parent on $\Pr(y_i = 1|\beta)$.

(e) Now assuming the prior distribution $\beta \sim N(0_k, I_k)$, estimate the model in Part (b) using the Bayesian method. Report the posterior estimates and posterior standard deviations from 20,000 Gibbs iterations after a burn-in of 5,000 iteration. Present the results in a table.

(f) Calculate the covariate effect of increasing the age by 5 years on $\Pr(y_i = 1|\beta)$, marginal of the parameters and the remaining covariates. Is your answer different compared to Part (c) of the question?

4. **(25 points)** Consider the following longitudinal data model or mixed effects model,

$$y_{it} = x_{it}'\beta + w_{it}'b_i + u_{it}, \qquad i = 1, \cdots, n;\ t = 1, \cdots, T, \tag{1}$$

where x_{it}' is $1 \times k_1$, w_{ij}' is $1 \times k_2$, β is $k_1 \times 1$, b_i is $k_2 \times 1$ and $u_{it}|h_u \sim N(0, h_u^{-1})$. Assume the following prior distributions, $\beta \sim N_{k_1}(0, 10 * I)$, $h_u \sim Ga(6/2, 3/2)$, $b_i|D \sim N_{k_2}(0, D)$ and $D^{-1} \sim W_{k_2}(\nu_0 = 6, D_0 = I)$. Here I is an identity matrix of appropriate dimension.

The aim of this exercise is to study the effect of union membership on wages using data from the National Longitudinal Survey of Youth and consists of 545 observations observed between 1980-1987, contained in the file '`Vella-Verbeek-Data.xlsx`'. The response variable is logarithm of wages, intercept and experience variable should be assigned to w_{it} (random effects) and the remaining variables including the union membership should be assigned to x_{it} (fixed effects). Report the posterior mean and posterior standard deviations of all the parameters based on 10,000 MCMC iteration after a burn-in of 2,500 iterations. What is the premium on union membership?

To know more about the application, read Section 4.4 and Section 10.3 from the book 'Introduction to Bayesian Econometrics,' by Edward Greenberg.