Econometrics II, ECO351, Semester I, 2025-26
Homework II (100 points)

Instructor: M.A. Rahman

Deadline: 6:00 pm, September 12, 2025 (in my office: Room 409, ESB-2).

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 top and one inch from left. Staple the sheets on the top-left.*
- *Please submit computational assignments (if any) and written answers together and in the correct order. Your answer script should directly address the questions, with all code included in the appendix. All questions in this assignment must be completed using MATLAB. Work done in any other software will NOT be accepted.*

---

1. **(5+10+10 = 25 points)** Consider a multinomial logit (MNL) model (or Greene's conditional logit model) written as follows:

$$\begin{aligned} U_{ij}^{*} &= x_{ij}'\beta + \varepsilon_{ij}, \\ y_i &= k, \quad \text{if} \quad U_{ik}^{*} > U_{ij}^{*}, \quad \forall j \neq k, \end{aligned} \tag{1}$$

where U_{ij}^{*} denotes the latent utility for individual i $(= 1, 2, \cdots, n)$ for alternative j $(= 1, 2, \cdots, J)$. As usual x_{ij} and β are covariate and parameter vectors of dimension $k \times 1$; and the error ε_{ij} follows a *Type-1 Extreme Value (EV Type 1) or "Gumbel" distribution.*

(a) For the above MNL model, show that $\Pr(y_i = j) = P_{ij} = \dfrac{\exp(x_{ij}'\beta)}{\sum_{k=1}^{J}\exp(x_{ik}'\beta)}$. Also, write down the likelihood function.

We know that the marginal effect of a given attribute l associated with option j on the choice probability for option j can be derived as,

$$\frac{\partial \Pr(y_i = j)}{\partial x_{ij,l}} = \frac{\partial P_{ij}}{\partial x_{ij,l}} = \beta_l P_{ij}(1 - P_{ij}). \tag{2}$$

Similarly, the effect of this change on some *other option*, say k, becomes,

$$\frac{\partial \Pr(y_i = k)}{\partial x_{ij,l}} = \frac{\partial P_{ik}}{\partial x_{ij,l}} = -\beta_l P_{ij} P_{ik}. \tag{3}$$

Lets suppose we have simple a MNL model with $J = 3$ options and only one covariate x. With one covariate, the third subscript l in x is redundant and $x_{ij,l} = x_{ij}$, but we will use it to maintain clarity. We have,

$$\begin{aligned}
P_{i1} &= \frac{\exp(\beta_{11} + \beta_2 x_{i1,2})}{\exp(\beta_{11} + \beta_2 x_{i1,2}) + \exp(\beta_{12} + \beta_2 x_{i2,2}) + \exp(\beta_{13} + \beta_2 x_{i3,2})} \\
P_{i2} &= \frac{\exp(\beta_{12} + \beta_2 x_{i2,2})}{\exp(\beta_{11} + \beta_2 x_{i1,2}) + \exp(\beta_{12} + \beta_2 x_{i2,2}) + \exp(\beta_{13} + \beta_2 x_{i3,2})} \\
P_{i3} &= \frac{\exp(\beta_{13} + \beta_2 x_{i3,2})}{\exp(\beta_{11} + \beta_2 x_{i1,2}) + \exp(\beta_{12} + \beta_2 x_{i2,2}) + \exp(\beta_{13} + \beta_2 x_{i3,2})}.
\end{aligned}$$

We choose the third option as the base category and so set $\beta_{13} = 0$. For this three options MNL model, answer the following questions.

(b) Find the covariate effect of $x_{i1,2}$ on alternative 1 and express it in the formulation as given in equation 2.

(c) Find the covariate effect of $x_{i1,2}$ on alternative 2 and 3, and express them as in equation 3.

2. (5+10+5+5 = 25 points) Suppose consumers have to choose between three types ($J = 3$) of soft drinks, say `Pepsi`, `7-Up`, and `Coke Classic`, available in 2 -liter bottles. Shoppers will visit their supermarkets and make a choice, based on prices (`price`) and other factors (`feature` and `display`). The data for this exercise is present in the file "`cola.xlsx`" and is in stacked form. There are 3 records per customer id. The first record is for Pepsi, the 2nd for 7-Up and the 3rd for Coke. A description of the variables is below.

`id`	customer id.
`choice`	equals 1, if a specific brand is chosen.
`price`	price of a 2-liter soda.
`feature`	equals 1 if its a featured item at the time of purchase, otherwise 0.
`display`	equals 1 if displayed at time of purchase, otherwise 0.

Let y_{i1}, y_{i2}, and y_{i3} be indicator variables that indicate the choice made by individual i. If Pepsi is chosen, then $(y_{i1} = 1, y_{i2} = 0, y_{i3} = 0)$; else if 7-Up is selected, then $(y_{i1} = 0, y_{i2} = 1, y_{i3} = 0)$; and if Cola Classic is opted, then $(y_{i1} = 0, y_{i2} = 0, y_{i3} = 1)$. This y_i vector is given by $\texttt{choice}_i$ variable. The price facing an individual i for brand j is $\texttt{price}_{ij}$. That is, the price of Pepsi, 7-Up, and Coke is potentially different for each customer who purchases soda. Remember, different customers can shop at different supermarkets and at different times. Variables like `price` are *individual* and *alternative-specific* because they vary from individual to individual and are different for each choice the consumer might make. This type of information is very different from *individual specific* variables, which do not change across alternatives.

Based on the above information and the $J = 3$ option MNL in Question 1, complete the following task.

(a) Suppose that we observe three individuals, who choose alternatives one, two, and three, respectively. Assuming that their choices are independent and we only include `price` as the independent variable, what is the probability of observing this outcome?

(b) Fit a multinomial logit model (Greene's conditional logit model) by regressing `choice` (our y variable) on `price` only (ignore `feature`, and `display`). Report the coefficient estimates, standard errors, and t-statistics.

(c) Compute the probabilities P_{i1}, P_{i2}, and P_{i3} with prices set as follows: $\texttt{price}_{i1} = \$1.0$, $\texttt{price}_{i2} = \$1.25$, and $\texttt{price}_{i3} = \$1.10$.

(d) With prices of soda set as mentioned above, utilize equations (2) and (3) to compute the marginal effects for a \$0.10 (i.e., 10 cent) increase on all three choices.

3. **(4+6+4+4+4+4+4 = 30 points)** The data file "`FMD.xlsx`" corresponds to number of incidents of foot and mouth disease (FMD) amongst cattle in different provinces of Turkey for 1998. The objective is to assess any difference in FMD incidence between the eastern and western regions of the country. Based on the above information, answer the following.

(a) Present a descriptive summary (mean, median, standard deviation, maximum, and minimum) of the variables "`FMD1998`" and "`Cattle`" classified according the two different regions of Turkey. What do you observe?

(b) Fit the following Poisson regression model,

$$\log(\lambda_i) = \beta_1 + \beta_2 \ln(\texttt{Cattle}_i) + \beta_3 \,\text{EasternTurkey}_i, \qquad \text{for all } i = 1, 2, \ldots, n,$$

where `EasternTurkey`=1, if the province lies in Eastern Turkey, and 0 otherwise. Report the MLE estimates of β, standard error, t-values, and the 95 percent confidence intervals.

(c) Interpret the coefficients β_2 and β_3 in the given context. Are these coefficients significant?

(d) What is the expected number of FMD in Eastern Turkey, when `Cattle`=170080?

(e) Calculate the probability that there will be 2 or less FMD in Eastern Turkey.

(f) What is the expected number of FMD in Western Turkey, when `Cattle`=157850?

(g) Calculate the probability that there will be 2 or less FMD in Western Turkey.

4. (10+10 = 20 points) Sometimes a count data set contains more zeros than one would expect from a Poisson density. One possibility of modeling these "excess zeros" is via a Zero Inflated Poisson (ZIP) specification. The ZIP model is a mixture distribution that combines a binary probability model with a Poisson density. It states that the observed count variable y is strictly zero with probability π, and follows a generic Poisson distribution with probability $(1-\pi)$. This leaves two interpretations for an observed "0" in the data. An "always zero" interpretation, if the "0" is generated by the binary component or a "just happens to be zero but could have been > 0" if the "0" is generated by the Poisson.

(a) Derive the overall probability of observing a zero outcome for this mixture distribution and show that it is larger than the probability of observing zero flowing from a Poisson distribution alone (assume that $0 < \pi < 1$ and that the Poisson density is not degenerate, i.e. no single count receives a full probability of one).

(b) Consider a sample of n observations generated by this mixture distribution. Write down the likelihood function for a single observation for a "zero" and for a "nonzero" outcome. Then write down the likelihood for the whole sample.