Data Analytics for Business

Homework 2 – Part 1

Censored Models

The data file for this exercise is "Mobile_data_usage.csv", which is a data set extracted from a real research project. In that research project, we collaborated with a mobile carrier and would like to investigate mobile data users' dynamic consumption behavior, that is, how much mobile data each individual user consumes every day throughout their monthly billing cycle, depending on the amount of monthly quota remaining at the beginning of each day and the number of days left. It has the following variables:

Data Description

Column Name	Variable Description
CustomerID	An artificial identifier for each customer
DataUse	The amount of mobile data a customer consumes on a particular day (MB)
Quota	The amount of monthly mobile data quota still remaining at the beginning of each day (MB)
Days	Day count, equal to the number of days remaining towards the end of the billing cycle

Import the data into R and perform the following analysis.

Tasks:

1. Our main focus is to investigate how users' daily data usage amount **DataUse** changes with the remaining quota **Quota** when controlling the number of days left **Days**. First, create a scatter plot to visualize the data. Set **DataUse** as the y-axis and **Quota** as the x-axis. Save the plot from the R output below. What pattern in the data requires special treatments beyond the usual linear regression?

2. Estimate a linear model by simply regressing **DataUse** on **Quota** and **Days**. Use the **lm()** function and display the R output for the summary of the regression results.

3. Next, estimate the following Tobit model:

$$DataUse_{it}^{*} = \beta_0 + \beta_1 \cdot Quota_{it} + \beta_2 \cdot Days_{it} + \varepsilon_{it}$$

$$DataUse_{it} = \begin{cases} DataUse_{it}^{*} & \text{if } DataUse_{it}^{*} \geq 0 \\ 0 & \text{if } DataUse_{it}^{*} < 0 \end{cases}$$

Use the **censReg()** function included in the **censReg** package. Include all the R output for the summary of the regression results.

4. List together and compare the estimates of the linear model and the Tobit model. (**Note**: only include the **first three elements**: intercept, Quota, Day.) How does the coefficient for **Quota** from the Tobit model compare to that from the linear model? Why is it greater/less?

5. Compute the **marginal effects** of **Quota** based on the Tobit model at two different X values: (i) **Quota = 10**, Days = the mean of Days in the sample; (ii) **Quota = 2,000**, Days = the mean of Days in the sample. Use the **margEff()** function in the **censReg** package. (**Hint**: don't forget to include 1 in the xValues argument to account for the intercept.)

   Compare the marginal effects with that of the linear model. (Note that the marginal effects of a linear model simply equal the coefficients.) What conclusions and implications can you draw? What do these marginal effect values mean? How and why do the marginal effects vary as the explanatory variables change?