**15.835 ENTREPRENEURIAL MARKETING**
**Assignment II: Demand Forecasting**

This assignment is designed to help students learn about how to forecast the demand of a new product or service. For this assignment, students should have an execution file, BLOGIT.EXE and a data file, DATA1.DAT. The data file contains *synthetically generated* choice observations and the execution file will implement the binary logit analysis. In the Appendix, students can find a concise description about how to use the execution file.

## Demand Forecasting of a New Product/Service (Segway Inc. )

On Dec. 3, 2001, Dean Kamen unveiled Segway Human Transporter (HT), the world's first dynamic self-balancing human transporter, which was designed to enhance the productivity of people by increasing the distance they can travel and the amount they can carry. The Segway HT is an one-person, gyroscope-packed, electric-powered scooter and perhaps the biggest innovation in transportation since the car replaced the horse and buggy.

Kamen and his team plan to release their products to customers by the end of this year. Their commercial ambitions are as advanced as their technical virtuosity. By stealing a slice of the $300 billion transportation industry, they expect that their start-up, Segway Co., will be the fastest outfit in history to reach $1 billion in sales. To get there, Kamen has collected a total of about $90 million and has erected a 77,000-sq-ft. factory a few miles from its Manchester, N.H., headquarters that will be capable of churning out 40,000 Segway HTs a month by the end of this year.

Before the market launch, Segway Co. should resolve two strategic issues: (1) which segment is the best initial target market? and (2) what is the best product for the chosen initial target market? Even though they plan to eventually penetrate both commercial and consumer markets, they feel they need to focus on only one market in the short term in order to generate the revenue necessary to upgrade their products. Since there is a significant difference between two markets in terms of needs, the technical specification of the first commercial transporters depends on the chosen market. In particular, Segway Co. is considering three attributes as follows:

Price : $3,000 or $8,000
Maximum Speed: 10mph or 20 mph
Range: 17 miles or 30 miles per battery charge

To address these two issues, a group of brilliant Sloan MBA students have conducted a market research study designed to measure purchase intents for alternative product concepts. The data they collected are given in a tab-delimited file, DATA1.DAT, which contains total 300 purchase intent observations, 150 from commercial customers and 150 from public consumers. The data file consists of 6 columns as follows:

Col. No. 1: ID (1-300)
Col. No. 2: Price (0 for $3,000; 1 for $8,000)
Col. No. 3: Maximum Speed (0 for 10 mph; 1 for 20 mph)
Col. No. 4: Range (0 for 17 miles; 1 for 30 miles)

Col. No. 5: Observed Purchase Intent (0 for no purchase intent; 1 for purchase intent)
Col. No. 6: Customer characteristics (0 for consumers; 1 for commercial customers)

Since purchase intents (Col. No. 5: CN5) are binary responses, the binary logit model is suitable for this data set. In particular, we assume the binary logit model as follows:

$$\text{Prob of customer } i\ (i = 1, \ldots, I) \text{ making a purchase} = p_i = p(\text{CN5}_i = 1) = \frac{e^{\beta' x_i}}{1 + e^{\beta' x_i}}, \quad (1)$$

where $x_i = (1, CN2, CN3, CN4)$ is a 4-dimensional column vector of an intercept and three independent variables (covariates) and $\beta' = (\beta_1, \beta_2, \beta_3, \beta_4)'$ is a 4-dimensional row vector of coefficients for $x_i$. Note that the first element of $x_i$, 1, is designed to capture an intercept and the first element of $\beta, \beta_1$, denotes the estimate of the intercept.

1. Fit the binary logit model to DATA1.DAT by using BLOGIT.EXE and
   a. Get the estimates of $\beta$ for all data,
   b. Get the estimates of $\beta$ only for commercial customers,
   c. Get the estimates of $\beta$ only for consumers.

   After finding three sets of $\beta$'s, compare them each other in terms of magnitude and discuss implications of this.

   Note: for this task, you should run BLOGIT.EXE three times with different options: sel=all for a, sel=(6:1) for b, sel=(6:0) for c.

2. Fill out the following table.

| Product Concept | Pred. M/S* in whole market | Pred. M/S in commercial market | Pred. M/S in consumer market |
|---|---|---|---|
| A ($3K, 10mph, 17miles) | | | |
| B ($3K, 10mph, 30miles) | | | |
| C ($3K, 20mph, 17miles) | | | |
| D ($3K, 20mph, 30miles) | | | |
| E ($8K, 10mph, 17miles) | | | |
| F ($8K, 10mph, 30miles) | | | |
| G ($8K, 20mph, 17miles) | | | |
| H ($8K, 20mph, 30miles) | | | |

Note: *: Predicted market share

Hint: You can use Eq. (1) to compute the predicted M/S since you have the estimate of $\beta$ and since $x_i$ can be defined for each product concept.
For example, $x_i$ becomes $(1, 0, 1, 0)$. and $(1, 0, 1, 1)$ for product C and D, respectively.

**(one more question in the next page)**

3. Discuss
    a. Should Segway focus on only one market?
    b. If Segway wants to focus on only one market in the short term, which market is the best initial target market?
    c. Which product concept is best for the chosen target market?
    d. To penetrate the other (unchosen) market in the future, what should Segway do?

# APPENDIX

- Before using BLOGIT.EXE
  1. Create a folder(directory) in your hard drive (e.g., c:\> md EMKTG)
  2. Copy BLOGIT.EXE and DATA1.DAT to the folder.

- How to use BLOGIT.EXE

1. BLOGIT.EXE is a simple execution file.
2. BLOGIT.EXE requires a program file which contains all parameters for options you should provide. The options you should provide are as follows:

data=*name_of_data_file*;
out=*name_of_ file_ for_output*;
save=*name_of_file_for_saving_predicted_choice_probabilities*;
nrow=*number_of_rows_in_data_file*;
ncol=*number_of_columns_in_data_file*;
id=*index_of_column_in_data_file_which_has_id*;
cov=*index_of_columns_in_data_file_which_will_be used_as_covariates*;
ch=*index_of_column_in_data_file_which_has_observed_choices*;
intercept=*indicator_for_intercept*;
sel=*indicator_for_observation_selection*;

Rule 1: There should be NO SPACE.
2: Each option line should be ended by ";".
3: File names can have a folder information.
4. For COV option, column numbers should be delimited by commas",".
5. For INTERCEPT option, a) y if you want to include an intercept
b) n otherwise.
6. For SEL option, a) all if you want to use all available observations in data file
b) (*Column_number_in_data_file*:*specific_integer_value*)
if you want to use a subset of observations in data file.
Note that you should provide ":" in the option field.
For example, if you give sel=(2:3),
a subset of observations in the data file will be
used only if its 2nd column has a value of 3.

Here is the program file you should use for DATA1.DAT .

```
data=c:\EMKTG\data1.dat;      /* Observations will be imported from
                                 c:\EMKTG\data1.dat */
out=c:\EMKTG\all.out;         /* Estimation result will be written into
                                 c:\EMKTG\all.out */
save=c:\EMKTG\all.sav;        /* Predicted choice probabilities will be written
                                 into c:\EMKTG\all.sav */
nrow=300;                     /* c:\EMKTG\data1.dat has 300 rows.*/
ncol=6;                       /* c:\EMKTG\data1.dat has 6 columns */
id=1;                         /* The 1st column in c:\EMKTG\data1.dat contains IDs */
cov=2,3,4;                    /* The 2nd, 3rd, and 4th columns in c:\EMKTG\data1.dat
                                 will be used as covariates */
ch=5;                         /* The 5th column in c:\EMKTG\data1.dat contains
                                 observed choice decisions */
intercept=y;                  /* An intercept will be included */
```

```
sel=all;                    /* All observations in c:\EMKTG\data1.dat will be used */
```

3. Use Notepad to type the above options and save them as a text file in the folder you created.
   (e.g., c:\EMKTG\all.prg)
4. To run BLOGIT.EXE, click Start->Programs->Accessories->Command Prompt
5. Then, type c:\> cd *the_folder_name you_created* (e.g., c:\>cd EMKTG)
6. Then, type blogit *file_name_of_your_program_file*
   e.g., c:\EMTG>blogit all.prg
7. The analysis will be done shortly. Examine output and save files.

- Example of output file

```
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+ Binary Logit Analysis
+ Copyright (c) 2002 by Jin Gyo Kim
+ Sloan Sch. of Mgmt, MIT
+ kimjg@mit.edu
+
+ This program was coded to supplement class discussions.
+ This program can be distributed freely only for educational purposes.
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

    Data File : eg.dat
        No. of rows : 300
        No. of columns : 10
        ID : Column No. 1
        Choice observation : Column No. 4

    3 Covariates
        Intercept and Col. No. 2, 3

    All 300 observations are used

Iteration   Log-likelihood
---------   --------------
    1       -97.8441
    2       -90.3836
    3       -89.7739
    4       -89.7609

    As log-likelihood value increases, it is better.

Total No. of Iteration: 4

Log-likelihood: -89.76090

Estimation Result  Estimate of β
    Variable    Estimated Beta    Std. Dev.    Z Value
    ---------   --------------    ---------    -------
    Intercept        2.88839       0.44149      6.5423
    Col. No. 2      -4.82683       0.45288    -10.6582
    Col. No. 3       0.12782       0.40127      0.3185

    The significance level can be determined by examining Z values.
    If |Z|>1.64, we may infer the estimate is significantly different from 0.
```

Classification Table

| Observed | Predicted 0 | Predicted 1 | Percentage correct |
|---|---|---|---|
| 0 | 137 | 7 | 95.14 |
| 1 | 21 | 135 | 86.54 |

Overall percentage: 90.67

Note: Predicted choice probabilities are saved in a file ex.sav
In the file, there are two columns.
1st col.: ID
2nd col.: Predicted Choice Prob.