Choice-Lab Marketing Case

Binary Logit with Customer Characteristics

Files to use with BIOGEME:
Model file: *BL_Marketing_mod_1.mod*
Data file: *marketing_case.dat*

In this model we try to assess what are the factors characterizing customers choice of dropping out as clients from Choice-Lab. The decision maker (Choice-Lab customer) faces a binary choice: either to remain as a client or drop as a client. The dependent variable (Choice) equal to 1 if the customer 'drops' next year and 0 otherwise. The model is estimated on the following variables:

- *NegProfit*: dummy variable for negative profit,
- *NegEquity*: dummy variable for negative equity,
- *LRSC*: dummy variable indicating if the legal status of the firm is limited responsibility stock owned company,
- *LnEmpl*: natural log of total number of employees, and
- *LnAge*: the natural log of the company's age.

For estimation purposes we normalize the alternative *remain client* and the estimated parameters are therefore interpreted relative to it. The following are the systematic part of the utilities for the two alternatives:

$$
\begin{aligned}
V_{\text{remain}} &= \text{ASC}_{\text{remain}} \\
V_{\text{drop}} &= \text{ASC}_{\text{drop}} + \beta_{\text{NegProfit}}\text{NegProfit} + \beta_{\text{NegEquity}}\text{NegEquity} + \\
&\quad \beta_{\text{LRSC}}\text{LRSC} + \beta_{\text{Empl}}\text{LnEmpl} + \beta_{\text{Age}}\text{LnAge}.
\end{aligned}
$$

```
[Choice]
Choice + 1

[Beta]
// Name   Value     LowerBound UpperBound  status (0=variable, 1=fixed)
ASC_remain      0.0          -100.0     100.0         1
ASC_drop        0.0          -100.0     100.0         0
b_NegProfit     0.0          -100.0     100.0         0
b_NegEquity     0.0          -100.0     100.0         0
b_LRSC          0.0          -100.0     100.0         0
b_Empl          0.0          -100.0     100.0         0
b_Age           0.0          -100.0     100.0         0

[Utilities]
// Id  Name  Avail  linear-in-parameter expression
1   Alt1  avail  ASC_remain * one
2   Alt2  avail  ASC_drop * one + b_NegProfit * NegProfit +
    b_NegEquity * NegEquity + b_LRSC * LRSC + b_Empl * LnNbEmpl +
        b_Age * LnAge

[Model]
// Currently, only MNL (multinomial logit), NL (nested logit), CNL
// (cross-nested logit) and $NGEV (Network GEV model) are valid keywords
//
$MNL

[Expressions]
// Define here arithmetic expressions for name that are not directly
// available from the data
one  = 1
avail = 1
```

Figure 1: Snapshot from the *BIOGEME* code

Figure 1 shows a snapshot of the BIOGEME code that corresponds to the systematic part of the utility functions. Section `[Choice]` indicates the dependent variable in the dataset, this is the variable identifying the chosen alternative (can be also a formula). Section `[Beta]` lists the parameters which we intend to use in our systematic utilities. If the status is set to 1 that means that the parameter is kept fixed at its value (zero, eg. for the ASC_remain), otherwise it is estimated. This is how we normalize one of the alternative specific constants (`ASC_remain`). The parameter names must be exactly the same as those expressed in the `[Utilities]` section (note that BIOGEME is case sensitive). In `[Utilities]`, we define the systematic utilities. Since both options are applicable for all customers, we have set the availability to be 1 in the `[Expressions]` section.

Estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	$\mathrm{ASC}_{\mathrm{drop}}$	-0.535	0.0880	-6.08
2	β_{LRSC}	-0.233	0.0470	-4.97
3	β_{Empl}	-0.185	0.0143	-12.97
4	β_{Age}	-0.097	0.0285	-3.40
5	$\beta_{\mathrm{NegEquity}}$	0.184	0.103	1.78
6	$\beta_{\mathrm{NegProfit}}$	0.198	0.0483	4.10
Summary statistics				
Number of observations $= 15934$				
$\mathcal{L}(0) = -11044.6$				
$\mathcal{L}(\hat{\beta}) = -7590.13$				
Likelihood ratio test $= 6908.95$				
$\bar{\rho}^2 = 0.312775$				

Table 1: Estimation results

The estimation results for this first model (BL_Marketing_mod_1.mod) are shown in table 1. In addition to the output listed above, BIOGEME provides further information in the output file; see BIOGEME user manual

for details.

Given our specification, the negative sign of ASC_{drop} mean that if the reminder of the utility is equal to zero, than a business is more likely to remain a customer. This can be interpreted as the decision maker prefers to remain client to the company.

The variable Age is negative and statistically significant indicating that the older the customer (age of the firm), the less likely it is to leave the company. Note that the variable Age also can capture other effects. Young firms might be more vulnerable to be closed down given financial difficulties (in need to cut costs), so this could explain why they decide to drop out. However there might also be other viable explanations, for example new firms might be interested in buying a one time list of addresses for direct marketing purposes (i.e. product 3).

The significant and negative coefficient of the parameter $LRSC$ (limited responsibility stock companies) implies that stock owned limited responsibility companies are less likely to drop as clients compared to non-stock limited responsibility firms.

The coefficient on the variables $Empl$ is negative and statistically significant and implies that larger firms are less likely to drop out. It could be that large firms are better established on the market, or may be operating in industries where access to companies' financial information is key for their success. This could be, for example, banks and financial institutions. We could also speculate that large companies have larger client databases and establish credit policies based on credit rating information provided by Choice-Lab. A small company might only buy one time credit rating report for one of its clients and this might happen very sporadically.

On the financial variables indicators, only negative profit is statistically significant different from zero. Companies needing to cut costs drop out as clients from Choice-Lab, as expected.

Binary Logit with Type of Purchased Product

Files to use with BIOGEME
Model file: *BL_Marketing_mod_2.mod*
Data file: *marketing_case.dat*

In this model we keep all the above independent variables, and add a set of variables describing the product purchased by the decision maker. The idea is to verify if there are any patterns of loyalty that can be explained by the type of products that clients have purchased. The systematic part of the utilities are:

$$
\begin{aligned}
V_{\text{remain}} &= \text{ASC}_{\text{remain}} \\
V_{\text{drop}} &= \text{ASC}_{\text{drop}} + \beta_{\text{NegProfit}}\text{NegProfit} + \beta_{\text{NegEquity}}\text{NegEqutiy} + \\
&\quad \beta_{\text{LRSC}}\text{LRSC} + \beta_{\text{Empl}}\text{LnNbEmpl} + \beta_{\text{Age}}\text{LnAge} + \\
&\quad \beta_{\text{IndAnalysis}}\text{IndAnalysis} + \beta_{\text{CreditInfo}}\text{CreditInfo} + \\
&\quad \beta_{\text{Accounts}}\text{Accounts} + \beta_{\text{Monitor}}\text{Monitor} + \\
&\quad \beta_{\text{Web}}\text{Web} + \beta_{\text{CD}}\text{CD} + \beta_{\text{CRM}}\text{CRM} + \beta_{\text{Internet}}\text{Internet} + \\
&\quad \beta_{\text{OpenDB}}\text{OpenDB} + \beta_{\text{Other}}\text{Other}.
\end{aligned}
$$

In table 2 we report the estimation results for this model. All product choice variables have a negative sign and are statistically significantly different from zero. However they vary in magnitude. The largest coefficients are found for the products that provide integrated and web based services (*CRM, Internet and Web*), which are the solutions that provide clients with the most complete and updated data. We could speculate that these might be solutions that clients use most frequently and that play an important role in their day to day decisions.

We have now identified some variables that have a significant impact on customer drop outs. However, we can provide Choice-Lab with an extra, valuable piece of information; a list of top 100 clients that have the highest probability of dropping out in the next year. Since we have data only until 2002, what we can calculate is the probability that a client will drop out in

Estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	ASC_{drop}	1.494	0.115	12.97
2	β_{LRSC}	-0.169	0.0497	-3.413
3	β_{Empl}	-0.131	0.0153	-8.546
4	β_{Age}	-0.216	0.0322	-6.739
5	$\beta_{NegEquity}$	0.321	0.113	2.83
6	$\beta_{NegProfit}$	0.274	0.0519	5.29
7	β_{CRM}	-2.556	0.707	-3.613
8	$\beta_{Internet}$	-2.244	0.139	-16.16
9	β_{Web}	-2.254	0.117	-19.26
10	β_{CD}	-1.798	0.070	-25.47
11	$\beta_{Monitor}$	-1.018	0.252	-4.038
12	$\beta_{IndAnalysis}$	-1.084	0.0624	-17.357
13	$\beta_{Accounts}$	-1.039	0.0635	-16.367
14	β_{Other}	-0.613	0.0538	-11.374
15	$\beta_{CreditInfo}$	-0.567	0.0540	-10.508

Summary statistics

Number of observations $= 15934$

$\mathcal{L}(0) = -11044.6$

$\mathcal{L}(\hat{\beta}) = -6718.4$

Likelihood ratio test $= 8652.42$

$\bar{\rho}^2 = 0.391$

Table 2: Estimation results

2003. One way of doing so is to divide the dataset in two samples: training sample and test sample. First we use the training sample (2000-2001), and estimate the model. Second, we calculate the predicted probability with the test sample (2002) using the model estimated from the training sample. Third, we list the data in descending order and pick the 100 clients with the largest probability. Choice-Lab could analyze the listing and decide for which clients is worth considering a retention strategy. We remind the reader that the dataset also includes other variables. Therefore it is advised to improve the specification and run additional models.