*International Review of Business Research Papers*
*Vol. 4 No.3 June 2008 Pp.281-290*

# An Application of Qualitative Choice Model to Estimate the Demand for Telecommunications Products

**Manzoor E. Chowdhury and Sonia H. Manzoor**

*Telecommunications is a growing industry in many developing countries including Bangladesh. With increasing wireless penetration, the telecom market in these countries will become more competitive, and companies will need more information on demand and market segments to efficiently utilize their marketing budget. This study identifies several socio-economic and product use characteristics of customers who are more likely to respond to a telemarketing campaign launched by a major telecommunications company in the United States. Given the discrete (binary) nature of the purchase decision (either buy or do not buy), this analysis relies on the use of Logistic Regression, one of the widely used techniques of qualitative choice modeling. The objective of the study is to demonstrate the application of a valuable tool in marketing research, particularly in the growing telecommunications industry.*

Field of Research: Market Research, Applied Econometrics

## 1. Introduction

This study identifies several socio-economic and product use specific characteristics of customers who are more likely to respond to a Message Line telemarketing campaign launched by a major telecommunications company in the United States. MessageLine is a network based voice mail system that allows one to conveniently retrieve and store messages from any touch-tone phone. Given the discrete (binary) nature of voice mail purchase decision (either order or do not order voice mail), this analysis relies on the use of Logistic Regression,

---

Manzoor E. Chowdhury, Independent University – Bangladesh (IUB)
Email: manzoorchowdhury@yahoo.com

Sonia H. Manzoor, Independent University – Bangladesh (IUB)
Email: soniahmanzoor@yahoo.com

one of the widely used techniques of qualitative choice modeling. Logit models, like discriminant analysis, focus on the relationship between group membership and a set of independent variables (predictors). But while discriminant analysis centers on the question "which group is the observation likely to belong to?" logit models focus more on estimating "how likely is the observation to belong to each group?"

## 2. Previous Literature

There is an extensive literature on logistic regression and its application. We have touched upon only a few studies and a more detailed related literature can be requested from the authors. Although logistic regression has been used in a variety of areas, for example in childhood ADHD context (Soldin et al. 2002), logistic regression has also been used in customer analysis. For example, Buckinx et al. have used logistic regression for partial detection of customers in retail setting (Buckinx et al., 2005). The method has also been used for predicting the customer's future profitability, based on his demographic information and buying history in the book club (Ahola et al., 2001). Hwang et al. (2004) and Mozer et al. (2000) used logistic regression for churn prediction in wireless communications industry. Crotts (2004) investigated the effect of cultural distance on overseas travel behavior using logistic regression.

## 3. Model and Data

This analysis centers on the hypotheses that several variables influence the decision to purchase voice mail service: CPE, CCF (single feature or package), network, CLASS (single feature or package), stand-alone features such as Call Waiting, Three-way Calling, Return Call, Repeat Dial, Caller ID, Total Voice, Signal Ring, and demographic variables such as income, occupation, age, education, and socio-economic segments. Table 1 discusses these predictor variables that were included in the model.

To more accurately predict the likelihood of Voice Mail purchase decision, the dependent variable Voice Mail included only those customers who have had voice mail for four or more months. A very high number of voice mail customers are expected to have the Call Forwarding feature. High correlation between variables results in unreliable regression coefficients. To isolate this correlation between Call Forwarding and Voice Mail, a few variables were modified where Call Forwarding feature was taken out from variables such as CCF, CCFPKG, Network etc. For example, a new variable called 'Non-Call Forwarding-CCF' was created where Call Forwarding feature was not included.

The statistical model for this study is given by:

$$L_n\,(p/1\text{-}p) = \text{constant} + \beta_1\,(\text{CPE}) + \beta_2\,(\text{NON-CF-CCF}) + \beta_3\,(\text{NON-CF-CCFPKG}) + \beta_4\,(\text{NON-CF-NETWORK}) + \beta_5\,(\text{NON-CF-ANYPKG}) + \beta_6\,(\text{CLASS}) + \beta_7\,(\text{CLASSPKG}) + \beta_8\,(\text{CW}) + \beta_9\,(\text{THREEWAY}) + \beta_{10}\,(\text{RC}) + \beta_{11}\,(\text{RD}) + \beta_{12}\,(\text{CID}) + \beta_{13}\,(\text{TV}) + \beta_{14}\,(\text{SR}) + \beta_{15}\,(\text{INCOME1}) + \ldots\ldots + \beta_{21}\,(\text{INCOME7}) + \beta_{22}\,(\text{OCCUP1}) + \ldots + \beta_{25}\,(\text{OCCUP4}) + \beta_{26}\,(\text{AGE1}) + \ldots. + \beta_{31}\,(\text{AGE6}) + \beta_{32}\,(\text{EDUC1}) + \ldots. + \beta_{36}\,(\text{EDUC5}) + \beta_{37}\,(\text{SEGMENT1}) + \ldots. + \beta_{44}\,(\text{SEGMENT8})$$

where the dependent variable is the log of odds (probability divided by one minus probability) whether a customer will buy VM or not, and the right hand side variables are the predictor variables that are discussed in Table 1. A customized random sample of 229,770 households was generated from the CRB database. The sample included only active residential customers to whom MessageLine voicemail was readily available. Column 3 in Table 2 shows the means of binary variables which reflect the proportions of customers that fall into a particular category. For example, in this sample roughly 5.4% of customers have CPE, about 5% have CLASS feature, about 11% are between 18 and 34 years of age, and about 5.3% of customers had a high school education.

## 4. Results and Implications

The maximum likelihood estimates of logit analysis are exhibited in Table 2. The regression required six iterations to generate the estimated parameter coefficients. Unlike linear regression where $R^2$ provides a measure of fit of the model, there is no such universally endorsed measure for logistic regression. One commonly used measure of goodness-of-fit involves the correct classification of customers as either ordering Voice Mail or not, on the basis of the information from the regressors. With a 50-50 classification scheme, approximately 94% of the observations were correctly classified as either trying or not trying Voice Mail, which indicates that the logit model performed quite well. This measure of goodness-of-fit involves the correct classification of decision-makers as either selecting the first alternative (yes) or the second alternative (no) solely on the basis of the explanatory variable information. Typically, if the estimated probability is greater than 0.5, then the first alternative is selected. On the other hand, the second alternative is selected if the estimated probability is less than 0.5. If the selected and actual outcomes match, the decision is correctly classified. If the predicted and actual outcomes do not conform as described, the decision is incorrectly classified.

The regression coefficients in Table 2 represent the change in the natural log odds of the dependent variable event (whether Voice Mail ordered or not) per unit increase in the corresponding independent variable. Since this does not provide sufficient intuitive explanation as well as does not provide any information on probabilities, changes in probabilities of the dependent variable due to a unit change in explanatory (predictor) variables are calculated and reported in the last

column of Table 2. Appendix I discusses the steps required to compute these probabilities.

**Table 1. Variable Definitions**

| | |
|---|---|
| CPE | 1 if household has CPE; 0 otherwise |
| NON-CF-CCF | 1 if household has non call forwarding CCF; 0 otherwise |
| NON-CF-CCFPKG | 1 if household has non call forwarding CCF package; 0 otherwise |
| NON-CF-NETWK | 1 if household has non call forwarding network; 0 otherwise |
| NON-ANYPKG | 1 if household has non call forwarding any package; 0 otherwise |
| CLASS | 1 if household has CLASS: 0 otherwise |
| CLASSPKG | 1 if household has CLASSPKG; 0 otherwise |
| CW | 1 if household has Call Waiting; 0 otherwise |
| THREEWAY | 1 if household has Three-way calling feature; 0 otherwise |
| RC | 1 if household has Return Call feature; 0 otherwise |
| RD | 1 if household has Repeat Dial; 0 otherwise |
| CID | 1 if household has Caller ID; 0 otherwise |
| TN* | 1 if household has Total Number; 0 otherwise |
| TV | 1 if household has total voice; 0 otherwise |
| SR | 1 if household has signal ring; 0 otherwise |
| INCOME1 | 1 if household has income between $10,000 and $19,999; 0 otherwise |
| INCOME2 | 1 if household has income between $20,000 and $29,999; 0 otherwise |
| INCOME3 | 1 if household has income between $30,000 and $42,499; 0 otherwise |
| INCOME4 | 1 if household has income between $42,500 and $62,499; 0 otherwise |
| INCOME5 | 1 if household has income between $62,500 and $87,499; 0 otherwise |
| INCOME6 | 1 if household has income between $87,500 and $149,999; 0 otherwise |
| INCOME7 | 1 if household income above $150,000; 0 otherwise |
| OCCUP1 | 1 if household occupation is homemaker, retired, or student; 0 otherwise |
| OCCUP2 | 1 if household occupation is Blue Collar; 0 otherwise |
| OCCUP3 | 1 if household occupation is White Collar; 0 otherwise |
| OCCUP4 | 1 if household occupation is Professional; 0 otherwise |
| AGE1 | 1 if household in the age bracket 18 to 24; 0 otherwise |
| AGE2 | 1 if household in the age bracket 25 to 34; 0 otherwise |
| AGE3 | 1 if household in the age bracket 35 to 44; 0 otherwise |
| AGE4 | 1 if household in the age bracket 45 to 54; 0 otherwise |
| AGE5 | 1 if household in the age bracket 55 to 64; 0 otherwise |
| AGE6 | 1 if household age is 65 or older; 0 otherwise |
| EDUCATION1 | 1 if household has some high school education; 0 otherwise |
| EDUCATION2 | 1 if household has high school education; 0 otherwise |
| EDUCATION3 | 1 if household has some college education; 0 otherwise |
| EDUCATION4 | 1 if household has college education; 0 otherwise |
| EDUCATION5 | 1 if household has graduate degree; 0 otherwise |
| SEGMENT1 | 1 if household belongs to socio-economic segment Buffs; 0 otherwise |
| SEGMENT1 | 1 if household belongs to socio-economic segment Young & Restless; 0 otherwise |
| SEGMENT1 | 1 if household belongs to socio-economic segment Dreamers; 0 otherwise |
| SEGMENT1 | 1 if household belongs to socio-economic segment Socialites; 0 otherwise |
| SEGMENT1 | 1 if household belongs to socio-economic segment Talkers; 0 otherwise |
| SEGMENT1 | 1 if household belongs to socio-economic segment Budget Conscious; 0 otherwise |
| SEGMENT1 | 1 if household belongs to socio-economic segment Self-Contained; 0 otherwise |
| SEGMENT1 | 1 if household belongs to socio-economic segment Golden Seniors; 0 otherwise |

* Variable excluded from the model because no household in the sample had Total Number.
** Description of these socio-economic segments is given in Appendix II.

**Table 2. Estimates of Logistic Regression**

| Variable | Coefficients | Mean | Change in Probability* |
|---|---|---|---|
| CPE | 0.63120 | 0.05367 | 0.08770 |
| NONCFCCF | 0.34680 | 0.07555 | 0.04820 |
| NONCCFPK | 0.11990 | 0.01807 | 0.01670 |
| NONNETWK | -0.24420 | -0.07754 | -0.03390 |
| NONANYPKG | -1.18370 | -0.17847 | -0.16450 |
| CLASS | -0.36870 | -0.05306 | -0.05120 |
| CLASSPKG | 0.10090 | 0.01257 | 0.01400 |
| CW | 0.60970 | 0.12933 | 0.08470 |
| THRWAY | 0.76160 | 0.00852 | 0.10590 |
| RC | 0.31660 | 0.01352 | 0.04400 |
| RD | 0.01450 | 0.00004 | 0.00200 |
| CID | 0.19990 | 0.02109 | 0.02780 |
| TV | 1.39640 | 0.00015 | 0.19410 |
| SR | 0.64910 | 0.00519 | 0.09020 |
| INCOME1 | -0.47610 | -0.02465 | -0.06620 |
| INCOME2 | -0.61330 | -0.04777 | -0.08520 |
| INCOME3 | -0.70800 | -0.11575 | -0.09840 |
| INCOME4 | -0.37470 | -0.06749 | -0.05210 |
| INCOME5 | -0.17950 | -0.01620 | -0.02490 |
| INCOME6 | -0.06040 | -0.00169 | -0.00840 |
| INCOME7 | 0.11380 | 0.00210 | 0.01580 |
| OCCUP1 | -0.35620 | -0.01048 | -0.04950 |
| OCCUP2 | 0.08900 | 0.00068 | 0.01240 |
| OCCUP3 | -0.03500 | -0.00110 | -0.00490 |
| OCCUP4 | -0.12200 | -0.00141 | -0.01700 |
| AGE1 | 0.34920 | 0.04289 | 0.04850 |
| AGE2 | 0.43860 | 0.06836 | 0.06100 |
| AGE3 | 0.27300 | 0.05166 | 0.03790 |
| AGE4 | 0.29250 | 0.02713 | 0.04070 |
| AGE5 | 0.06950 | 0.00590 | 0.00970 |
| AGE6 | -0.59430 | -0.08636 | -0.08260 |
| EDU1 | -0.27750 | -0.00044 | -0.03860 |
| EDU2 | -0.38510 | -0.05368 | -0.05350 |
| EDU3 | -0.08450 | -0.00074 | -0.01170 |
| EDU4 | 0.02390 | 0.00032 | 0.00330 |
| EDU5 | 0.04880 | 0.00015 | 0.00680 |
| SEG1 | 0.13770 | 0.01416 | 0.01910 |
| SEG2 | 0.01120 | 0.00012 | 0.00160 |
| SEG3 | -0.12310 | -0.01067 | -0.01710 |
| SEG4 | 0.16130 | 0.02089 | 0.02240 |
| SEG5 | 0.15610 | 0.01856 | 0.02170 |
| SEG6 | -0.25810 | -0.08126 | -0.03590 |
| SEG7 | -0.39410 | -0.03453 | -0.05480 |
| SEG8 | 0.12170 | 0.00550 | 0.01690 |
| Constant | -1.31810 | | |

* The entries in this column are equal to the product of the parameter estimates times the value of the standard normal probability density function. Appendix I discusses the steps required to compute the values in this column, i.e., changes in probability due to an unit change in independent variables.

The first coefficient on column 2 of Table 2 shows that customers with CPE (Customer Premises Equipment) are slightly more likely (probability is about 9%) to have voice mail compared to customers who do not have CPE. Although the change in probability is not considerably high, the CPE coefficient is the most statistically significant with the highest Wald statistic. Customers with CCF are slightly more likely (about 5% probability) to have voice mail compared to customers who do not have CCF, although CCFPKG came out as a weak predictor of voice mail purchase decision. Customers who have network are slightly less likely to buy voice mail compared to customers who do not have network (statistically significant with 3.39% lower probability). ANYPKG is a weak indicator of having VM with low statistical significance. Customers with CLASS feature are less likely to have voice mail, while customers with CLASSPKG are more likely to have VM. Among stand-alone features (CW, THREEWAY, RC, RD, CID, TV, and SR) all had positive signs, and with the exception of repeat dial and caller ID, all of them were statistically significant indicating that customers who have call waiting, three-way calling, return call, total voice, and signal ring are more likely to have VM as opposed to customers who do not have these features available to them. The changes in probabilities were highest for total voice, three way calling, and signal ring (19%, 10%, and 9%, respectively).

All income coefficients except INCOME7 had negative signs and only the coefficients of INCOME1, INCOME2, and INCOME3 are statistically significant. It can be concluded that although there was no significant relationship between VM and INCOME4 through INCOME7 (income between $42,500 and $150,000 plus), relatively low income customers (with income up to $42,499) are less likely to respond to a MessageLine voice mail campaign. The probabilities are also highest for INCOME2 and INCOME3 variables (8.5% and about 10% respectively).

The coefficient of OCCUPATION1 (homemaker, retired, and student) is negative and is also highly significant suggesting that people in this occupation are less likely to order messageline. Because customers in OCCUPATION1 fall in lower income bracket, this result is in agreement with results from income variables. However, no significant relationship was found for other occupation groups. The coefficients of AGE2 and AGE6 are statistically significant where AGE6 also had a negative sign. This suggests that customers who are most likely to respond to a message line campaign are relatively young (25 to 35 years of age) and customers who are 65 or older are less likely to respond. No significant relationships were found between education and voice mail purchase decision. Only the coefficient of EDUC2 (high school graduate) is negative and statistically significant implying that high school graduates are less likely to order message line. In agreement with *a priori* expectations, Segment1 (BUFFS), Segment4 (Socialites) and Segment5 (Talkers) are more likely to use VM compared to other segments. The negative signs and strong statistical significance of Segment6

and Segment7 suggest that Budget Conscious and Self-Contained segments are less likely to use voice mail.

It must be noted that no single variable stands out in predicting message line purchase decision. In addition, it is reported in Appendix I that the overall probability that message line will be ordered during the campaign is only about 16.7%. This may be partly due to the fact that in our sample (as well as in the entire database) only about 5% of the customers have had VM for four or more months. This may illustrate that prediction can be somewhat difficult in a sample with small number of observations where voice mail purchase decision was 'yes.'

## 5. Conclusion

This paper is a preliminary investigation to assist identifying target groups that are more likely to order message line voice mail service from a telecommunication company in the United States. To evaluate the goodness-of-fit of the logistic model, discriminant analysis was used on the same sample. The coefficients of discriminant analysis had similar signs and statistical significance, thus confirming the relative reliability of the logistic model. To keep the analysis simple and manageable, a number of variables were not included in this analysis. For example, the inclusion of three variables related to charges (interlata, intralata, and local) would have involved additional twenty one categorical variables for various ranges of charges.

The study is an example of how logistical regression can be used to help marketing campaigns and to reduce cost by appropriately targeting the customers who are more likely to purchase a particular product or service. Bangladesh has an emerging and growing telecommunications market where there is an urgent need for marketing research to understand customer behavior and demand. Using categorical data (through their own survey or database), the telecommunications companies can use tools such as logistic regression (and similar methods) to have more information about their customers.

## References

Ahola J., Rinta-Runsala E. 2001. "Data mining case studies in customer profiling," Research Report TTE1-2001-29, *VTT Information Technology*.

Buckinx W., Van den Poel D. 2005. "Customer base analysis: partial detection of behaviorally loyal clients in a non-contractual FMCG retail setting," *European Journal of Operational Research* 164, pp. 252-268.

Crotts, J. 2004. "The effect of cultural distance on overseas travel behaviors," Journal *of Travel Research* 43 August, pp. 83-88

Hwang H., Jung T., Suh E. 2004. "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunications industry. *Expert Systems with Applications* 26, pp. 181-188.

Mozer M. C., Wolniewicz R., Grimes D. B., Johnson E., Kaushansky H. 2000. "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," *IEEE Transactions on Neural Networks*, Special issue on Data Mining and Knowledge Representations.

Soldin O., Nandedkar A., Japal K., Stein M., Mosee S., Magrab P., Lai S., Lamm S. 2002. "Newborn thyroxine levels and childhood ADHD," *Clinical Biochemistry* 35, pp. 131-136.

## APPENDIX I

The Logistic Regression model can be written as:

$L_n\ (p/1\text{-}p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots. + \beta_k X_k$

This can be written as:

$P_i = F(Z_i) = \exp\ (Z_i)/1 + \exp\ (Z_i)$

Where $Z_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots. + \beta_k X_k$. $F(Z_i)$ is the (cumulative) logistic distribution function. The overall probability that Message Line will be ordered (based on this sample) can be calculated by computing $Z_i$ (at the sample means) as follows:

$Z_i$ = constant + coefficient of CPE * sample mean of CPE or percent of households with CPE +c coefficient of NON-CF-CCF * sample mean of NON-CF-CCF + ….. + coefficient of Segment 8 * sample mean of Segment 8

Or

$Z_i$ = -1.31810 + 0.63120 (0.05367) + 0.34680 (0.07555) + …… + 0.12170 (0.00550) = -1.60831

$F(Z_i) = \exp\ (Z_i)/1 + \exp\ (Z_i) = 0.1668$

Thus, the overall probability that Message Line will be ordered during the campaign is about 16.7%.

<u>Calculating change in probability due to a unit change in a predictor variable:</u>

Change in probability $P_i$ with respect to a change in each predictor variable $X_i$ can be calculated by multiplying the value of standard normal density function $f(Z_i)$ with each parameter coefficients $\beta_i$

$f(Z_i) = \exp\ (-1.60831)/[1 + \exp\ (-1.60831)]^2 = 0.13899$
$f(Z_i)$ * coefficient of CPE = 0.13899 * 0.63120 = 0.0877. Intuitively this means that a new customer with CPE is 8.8% more likely to order Message Line compared to a customer who does not have CPE. Similarly the other probabilities can be calculated by multiplying $f(Z_i)$ with respective parameter coefficients.

## APPENDIX II

Description of each socio-economic segment:

Segment 1 (BUFFS): Affluent younger customers with a high interest in new products and services.

Segment 2 (SOCIALITES): Affluent families with high usage and high social orientation.

Segment 3 (TALKERS): Middle-aged working and lower income families with high usage, a moderate social orientation, and a high price sensitivity.

Segment 4 (DREAMERS): Younger working and lower income customers with high interest in new products and services and high price sensitivity.

Segment 5 (SELF-CONTAINED): Affluent households with low price sensitivity and low interest and low usage in telephone products and services.

Segment 6 (YOUNG & RESTLESS): Young singles and couples with average to low interest in new products and services, high price sensitivity, and average to low usage.

Segment 7 (GOLDEN SENIORS): Affluent seniors with active lifestyles, above average usage and interest (for their age group) in products and services.

Segment 8 (BUDGET CONSCIOUS): Families and older singles and couples with very low income, low usage, and a low interest in new products.