2007

Predicting customer responses to direct marketing : a Bayesian approach

Wei CHEN

PREDICTING CUSTOMER RESPONSES TO DIRECT MARKETING: A BAYESIAN APPROACH

CHEN WEI

MPHIL

LINGNAN UNIVERSITY

2007

PREDICTING CUSTOMER RESPONSES TO DIRECT MARKETING: A BAYESIAN APPROACH

by
CHEN Wei

A thesis
submitted in partial fulfillment
of the requirements for the Degree of
Master of Philosophy in Business

Lingnan University

2007

ABSTRACT

Predicting Customer Responses to Direct Marketing:
A Bayesian Approach

By

CHEN Wei

Master of Philosophy

Direct marketing problems have been intensively reviewed in the marketing literature recently, such as purchase frequency and time, sales profit, and brand choices. However, modeling the customer response, which is an important issue in direct marketing research, remains a significant challenge. This thesis is an empirical study of predicting customer response to direct marketing and applies a Bayesian approach, including the Bayesian Binary Regression (BBR) and the Hierarchical Bayes (HB). Other classical methods, such as Logistic Regression and Latent Class Analysis (LCA), have been conducted for the purpose of comparison. The results of comparing the performance of all these techniques suggest that the Bayesian methods are more appropriate in predicting direct marketing customer responses. Specifically, when customers are analyzed as a whole group, the Bayesian Binary Regression (BBR) has greater predictive accuracy than Logistic Regression. When we consider customer heterogeneity, the Hierarchical Bayes (HB) models, which use demographic and geographic variables for clustering, do not match the performance of Latent Class Analysis (LCA). Further analyses indicate that when latent variables are used for clustering, the Hierarchical Bayes (HB) approach has the highest predictive accuracy.

Key Words: Direct Marketing; Bayesian Methods; Bayesian Binary Regression; Hierarchical Bayes; Logistic Regression; Latent Class Analysis.

DECLARATION

I declare that this is an original work based primarily on my own research, and I warrant that all citations of previous research, published or unpublished, have been duly acknowledged.

CHEN Wei
July 5th, 2007

CERTIFICATE OF APPROVAL OF THESIS

PREDICTING CUSTOMER RESPONSES TO DIRECT MARKETING: A BAYESIAN APPROACH

by

CHEN Wei

Master of Philosophy

Panel of Examiners:

____________________________ (Chairman)
Prof. Daning SUN

____________________________ (External Member)
Dr. Wenyu DOU

____________________________ (Internal Member)
Dr. Man-leung WONG

____________________________ (Internal Member)
Prof. Geng CUI

Chief Supervisor: Prof. Geng CUI

Co-supervisor: Prof. Tsang-sing CHAN

Approved for the Senate:

____
Prof. Mee-kau, NYAW
Chairman, Research and Postgraduate Studies Committee

____
Date

TABLE OF CONTENTS

CHAPTERS

LIST OF TABLES

LIST OF FIGURES

Acknowledgement

I am especially grateful to my Chief-Supervisor, Prof. Geng CUI, and Co-Supervisor, Prof. Tsang-sing CHAN, for their thorough guidance on my thesis. I also appreciate the financial support provided by Lingnan University, which makes it possible for me to complete my studies. Furthermore, I wish to extend my appreciation to Dr. Lianxi ZHOU, Dr. Patrick POON, Prof. Daning SUN, and other faculty members for their constructive advice on certain aspects of my thesis, Mr. Jun DU and Ms. Yuanyuan GUO for their research assistance, and Ms. Linda JO, Ms. Vanessa CHAN, and Ms. Daisy TSEUNG for their kindly help during my studies. Finally, I want to thank my parents for their love and encouragement.

Predicting Customer Responses to Direct Marketing: A Bayesian Approach

1 Introduction

Direct Marketing is a sales and promotion process in which the promotional materials and information are sent to individual customers via direct calling, mail, catalogue and so on (Bitran, and Mondschein, 1996). Nowadays, direct marketing has become more and more important in marketing practice. Not only a large portion of advertising expenditures are spent on direct marketing promotions, direct marketing sales also have a rapid growth rate.

Why has direct marketing become so important? One of the reasons is that, the cost of direct marketing is much lower than that of traditional methods, such as TV and newspaper advertisements. Another possible explanation is that, through callings, direct mails or catalogues, firms can build up their relationships with customers over time, and a good operation of such relationships may finally enhance the customer loyalty and improve profitability. Moreover, direct marketing firms can record individual customer information through the transactions, and it is likely that proper analysis of such information can lead to a better understanding of the customers and help firms to identify the potential target customers.

1.1 Direct Marketing Problems

Direct marketing differs from other marketing contexts in that it requires decision makers to make specific strategies toward individual customers. Both practitioners and researchers collect relevant customers' information, such as their transaction history and demographics, apply some techniques to analyze these data, and make decisions or conclusions according to their findings. In academic marketing literature, topics such as purchase frequency and time, sales profit, and brand choice are the direct marketing problems that are intensively examined. However, the predicting customer response, which is an important issue in direct marketing research, remains a significant challenge.

The prediction of customer response mainly focuses on identifying the potential buyers who can be called target customers or respondents. Here the term "target customers" or "respondents" refer to those customers who are the most likely to purchase a product if they are sent the promotional material or information. The number of respondents over the total number of mails sent is the called "response rate." As far as we know, the response rate of direct marketing promotions remains very low. According to the Response Rate Report of Direct Marketing Association (2005), the average response rate was just 5% or even lower in 2005. That is, if a firm sends promotion materials to 100 households, only 5 of them will eventually buy the products. A possible explanation for the low response rate is that direct marketing firms do not have adequate information about their customers and fail to make accurate prediction of customer behaviors. Nowadays many direct marketing firms not only record customer transactions but also purchase some second hand data such as the zip code level demographic information of customers.

Another possible explanation for the low response rate is the diversification of customer preferences and utilities. Methods that do not take this customer heterogeneity into account often make biased estimations. Researchers and practitioners attempt to classify customers into groups and analyze them separately. However, the reason why customers response to the promotion is not well understood because whether to response or not seems to have no direct relationship with variables such as demographic and geographic.

The development of direct marketing database makes it possible to derive more meaningful conclusions, but it also brings some problems. Since the constantly updating direct marketing database contains a large number of independent variables and observations, some classical techniques can not handle it properly due to the conflict of assumptions and computational deficiency.

Because of the budget limitation, it is impossible to send everyone in the population the promotional material. However, when only a few people are given direct mails or callings, the missing of potential customers will bring the company even more economic losses. So, it is crucial to build good predictive models to identify the target customers.

1.2 A Bayesian Approach

In order to solve the problems we mentioned above, we apply a Bayesian approach in this study for the customer response prediction using an empirical direct marketing data, and the performances are compared with those of other classical techniques including logistic regression and latent class analysis. This is also the main objective of this study.

First, in addition to the existing data, Bayesian methods can incorporate the prior information, which can be obtained from historical data and experts' opinion, into the modeling inference process. This is especially important in cases where insufficient data is available. In direct marketing context, the historical data such as transaction record provides us with opportunities to access the prior knowledge.

Second, the Bayesian methods are capable of analyzing the customer heterogeneity. For example, the Hierarchical Bayes assumes both the between- and within-group customer heterogeneity, which is often the case in practice, and makes individual customer level estimations. Hierarchical Bayes model is well matched with the need of direct marketing response prediction, where the decision is made individually.

Third, the Bayesian methods are free from the assumptions which are encountered by classical methods and are capable of analyzing the large and often poorly constructed direct marketing data.

Our results show that when analyzing customers as a whole group, the Bayesian Binary Regression has higher predictive accuracy than logistic regression. We also conduct the Hierarchical Bayes models using three demographic variables including customer type, economic status and state level geographic location, and discover that Hierarchical Bayes using these demographic variables has lower predictive accuracy than Latent Class Analysis. We further apply the Hierarchical Bayes using the same latent group membership as the Latent Class Analysis and notice the Hierarchical Bayes using latent variables has the best model performance. Overall speaking, our study indicates that the Bayesian methods can build more accurate models than other classical techniques such as logistic regression and Latent Class Analysis under the same circumstances and this indicates that the Bayesian methods are appropriate in direct marketing customer response prediction.

1.3 Organization of Study

The rest of the thesis is organized as follows: First, we review the literature of direct marketing modeling and the frequently used techniques, and some Bayesian applications in direct marketing. Second, we discuss the features of the Bayesian statistics and the Bayesian methods we use in this study. Third, we apply the Bayesian methods as well as other classical techniques using an empirical direct marketing database, and the results of customer response are compared. Finally, implications, limitations and future research directions are discussed.

2 Literature Review

No matter in what specific direct marketing situations, practitioners and researchers always want to identify the target customers that can bring them maximum profit through appropriate contact. Customers can be classified according to their purchase frequency, purchase amount, brand choice or preferences. In our case, direct mails of promotion materials are sent to customers and a binary dependent variable is measured. We use a direct marketing data and estimate the probability of response to the promotion for every single customer. Customers with the highest probability of response are considered the target customers.

However, we have to realize that every direct marketing firm has a large number of customers, normally several millions or even more. So it is inappropriate and often impossible to include all customers in the analysis. What practitioners and researchers do is to select samples from the dataset and apply a certain method to build the model. In this section, we first review some of the classical direct marketing techniques that are frequently used, and then discuss the advantages of the Bayesian methods.

2.1 Classical Direct Marketing Techniques

The most frequently used classical direct marketing technique is the so-called RFM model, which includes the Recency, Frequency, and Monetary value of customers' purchases (Colombo and Jiang, 1999; Rossi, McCulloch and Allenby, 1996; Gonul and Hofstede, 2006). Other popular methods include 1) Logistic Regression (Berger and Magliozzi, 1992), 2) Chi-square Automatic Interaction Detection or CH-AID (Bult, and Wansbeek, 1995; Haughton and Oulabi, 1997), 3)

Classification and Regression Trees, which is also referred to as CART, (Haughton and Oulabi, 1997), 4) Latent Class Analysis (LCA) (Jain, Bass and Chen, 1990), and 5) Artificial Neural Networks (ANNs) (Zahavi and Levin, 1997; Baesens, et al., 2002).

The term “classical” does not mean these techniques are out of date. Instead, we just use this term to distinguish these techniques from the Bayesian methods we apply in this study. Among those classical techniques, the Logistic Regression and the Latent Class Analysis (LCA) are applied in this study.

2.1.1 The RFM Model

In the past, the most commonly used method is the Recency, Frequency and Monetary value (RFM) model (Berger, and Magliozzi, 1992). Recency here means the number of consecutive mailings without response and the time period since the last order, while Frequency and Monetary value indicate the number of purchases made and amount of money spent in a certain time period respectively. In the modeling process, the RFM variables are divided into different categories, for instance, we may split the Recency variable into three categories: customers with purchases within the last, namely m days, between m and n days, and longer than n days (n>>m). Then we build contingency tables and estimate the probabilities for each category of group or individual customer according to their past response behaviors (Colombo and Jiang, 1999). Finally, we can use these probabilities to segment the customers and choose the target ones to make the offers.

The RFM method is very simple. It does not require any statistical software and the results can be easily understood and interpreted. But it also has some disadvantages. One is that the number of variables used may not enough because in real world, there are usually more variables that have effects on the probabilities of response, for instance, the demographic variables. The other is that, since RFM variables are not necessarily mutually independent, RFM model may have the problem of double counting (Bult and Wansbeek, 1995). Also, since the segmentation criteria are subjectively chosen, it may lead to overemphasis to the most attractive RFM segments and neglect other segments that would be profitable.

2.1.2 Logistic Regression

Another very commonly used technique is the Logistic Regression. The response variable of logistic regression is discrete or categorical, which makes it especially useful in building response model. Since the response variable can just have a value of 1 or 0, it may violate the assumption of the regression techniques that all response variables are normally distributed. Thus logistic regression assumes that the response variable does not necessarily conform a normal distribution, it can also distributed within the exponential family, like Poisson, binominal, and so on (Shepard, 1999).

The logistic regression can be written as a linear function of the predictor variables:

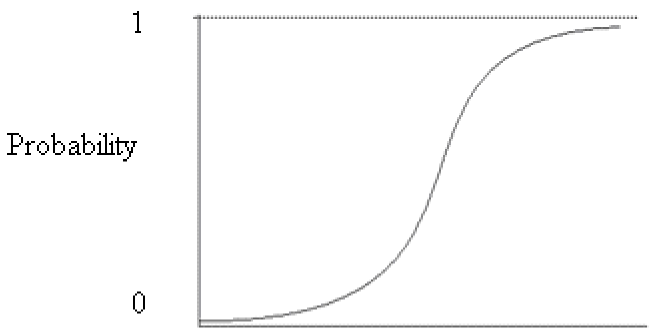
$$\ln[p/(1-p)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \tag{2.1.1}$$

$$\text{or } p = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k}} \tag{2.1.2}$$

In formula 2.1.1, p equals to the probability of an event occurring, p/(1-p) equals to the odds of an event occurring, and ln[p/(1-p)] equals to the log of the odds or logit. Here the occurring event indicates that the customers response to the mail or call.

First, the values of the predictor variables are used to form the likelihood function, and then, coefficients of all predictor variables are calculated by using Maximum Likelihood Estimation (MLE) iterations (Linder, Geier and Kolliker, 2004). After that, we can assign probabilities of response to each individual customer according to the parameters of the model. Figure 1 illustrates the relationship between dependent variable and independent variable in logistic regression.

Figure 1. Logistic Regression



The significance test of the influence of predictor variables can be accomplished using a Wald statistic test. In order to assess the model fit, a likelihood ratio test and Chi-square Goodness of fit are also conducted to compare the current model with the null model and the saturated model respectively.

Based on a statistical distribution, the Logistic Regression models generated are generally very robust (Rud, 2001). Also, the probabilities of each individual can be explicitly given and the regression coefficients can be easily understood and interpreted.

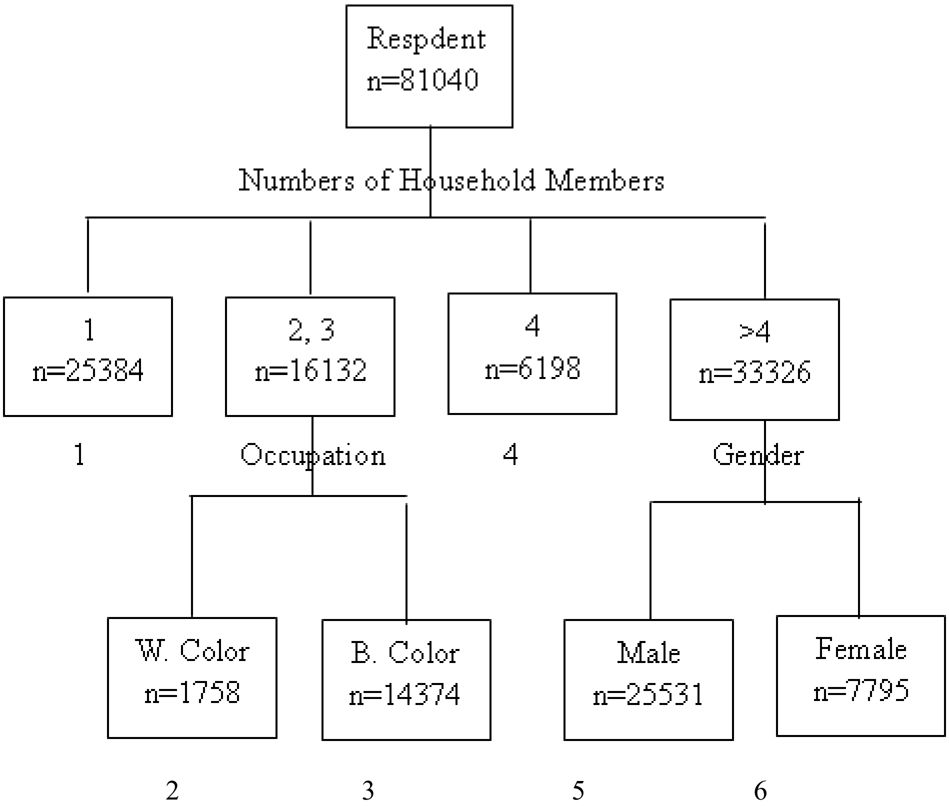
However, since direct marketing databases contain more and more independent variables and constantly update with new information, using Logistic Regression may cause some problems. First, because of the large number of predictor variables, using Maximum Likelihood Estimation (MLE) may fail due to the lack of convergence, large estimated coefficient variances, poor predict accuracy, and reduced power for testing hypotheses concerning model assessment (Genkin, Lewis and Madigan, 2005). Second, even if such numerical problems can be avoided, the asymptotic properties of maximum likelihood will sometimes result in an overfit of the data (Genkin, Lewis and Madigan, 2005). Third, majority of the predictor variables will be neglected during the variable selection process, which can improve the computational efficiency and human interpretation. However, according to Genkin et al. (2005), the unfounded statistical foundation of variable selection will make it difficult to choose the number of variables in a given task. Because Logistic Regression considers variables independently, redundant or ineffective combinations of variables may be chosen.

2.1.3 Chi-square Automatic Interaction Detection (CH-AID)

Chi-square Automatic Interaction Detection (CH-AID) was developed as an extension of RFM model by Kass in 1976. In this technique, all variables, both independent and dependent ones, are categorical. An essential feature of CH-AID is the use of the chi-square test for contingency tables to decide which variables are most important for classification. In this way observations will be split in order to make a minimal within group variance of the response. Variables with the lowest within group variance are selected and subdivided (Haughton and

Oulabi, 1997). The sub lists are analyzed in the same way. The whole CH-AID model is a tree-like structure and this technique can avoid the double counting problem of the RFM model (Bult and Wansbeek, 1995). Figure 2 gives a simple example of how CH-AID performs.

Figure 2. Chi-square Automatic Interaction Detection



An important feature of CH-AID is its ability of building non-binary classification trees, that is, trees where more than two branches may go from a node. Also, the CH-AID can deal with multi-way contingency tables, when both predictor and response variables have many classes. For this reason CHAID is widely used in the market segmentation. Another advantage of CH-AID is that it can discover the interaction variables that can not be identified by Logistic Regression (Shepard, 1999), but it does not produce the results similar as regression analysis.

Therefore, the output of CH-AID analysis could be applied to other methods such as Logistic Regression, and then build a model with more predictive power (Haughton and Oulabi, 1997).

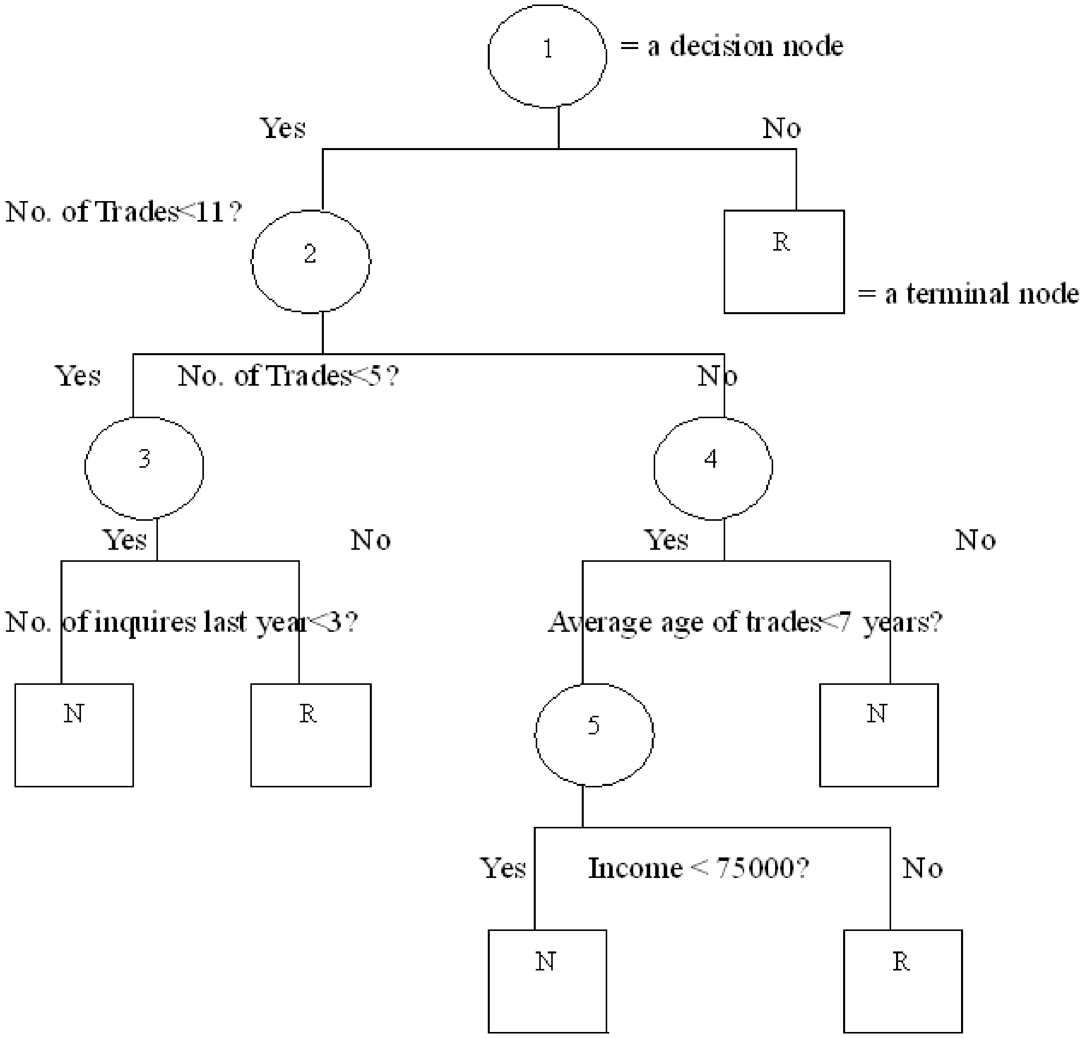
The CH-AID method also has some limitations. First, it can not handle a large number of independent variables. Nowadays, the CH-AID software is set to handle a maximum of 40 variables, which is inadequate to analyze the direct marketing data where the number of variables is far more than 40. Second, the CH-AID method requires all variables to be categorical. The users have to make a prior discretization process for those continuous predictor variables, which may result in a loss of information. Third, CH-AID method uses chi-square criterion to split observations. The splitting is stopped when the chi-squares are not significant. This splitting rule was highly criticized and is the main reason why CH-AID was not been widely accepted (Thrasher, 1991).

2.1.4 Classification and Regression Trees (CART)

Classification and Regression Trees (CART) is another tree building procedure. The whole process starts with an initial node and continues to separate the observations into decision nodes and terminal nodes. Here the decision nodes refer to the nodes that still need to be segmented while terminal nodes does not. The perfect split will classify all respondents to one direction and all nonrespondents to the other, but this rarely happens. Thus two criteria are set up to guide the splitting, called the Gini criterion and the twoing criterion. The Gini criterion is to find out the largest group in the data and try to isolate it with other groups. The towing criterion is to separate the data so that each group has equivalent observations. The splitting process stops when each observation is classified into one terminal node. Obviously, the tree size will grow very large since almost every direct marketing database contains a large number of observations. Because of this problem, the CART uses the following process to choose the tree size. First, the CART assigns few observations into each terminal node. Then, a tree pruning algorithm will be used to choose the appropriate tree size. That is, some of the branches will be deleted until the remaining tree has the smallest misclassification rate than those subtrees with the same size (Thrasher, 1991). Figure 3 illustrates the tree growing process.

CART is very useful in identifying segments with a desired behavior, such as response. Although CART has a very similar tree growing process to that of CH-AID, their underlying algorithms are totally different. The CH-AID uses the chi-square statistics to determine the tree size while CART uses the tree pruning algorithm. And the latter was proved to be more reliable (Thrasher, 1991). Unlike CH-AID, the CART does not have the limit of the number of variables, and it can analyze both categorical and continuous predictor variables. CH-AID can make multiple splits while CART can just make a binary one. However, according to Thrasher (1991), CART can split the same variable more then once, and successive splits using the same variable are equivalent to multiple splits. And these repeating splits will certainly obtain more information of the same variable. Like CH-AID, CART also gives insight into the interactions among variables. Both the tree processes are easy to use and interpret. Another important feature of CART is that it does not have any assumptions about the distribution of the predictor variables, which makes it more applicable.

Figure 3. Classification and Regression Trees



CART also has some weaknesses, one of which is that it is not based on a probabilistic model. There is no probability level or confidence interval associated with predictions derived from using a CART tree to classify a new set of data. Although the CART is considered to be more suitable for problem solving than CH-AID (Thrasher, 1991), Haughton and Oulabi's (1997) study results suggest that CART and CH-AID have similar response lift values, and indicate that the CART also has the potential of over fitting.

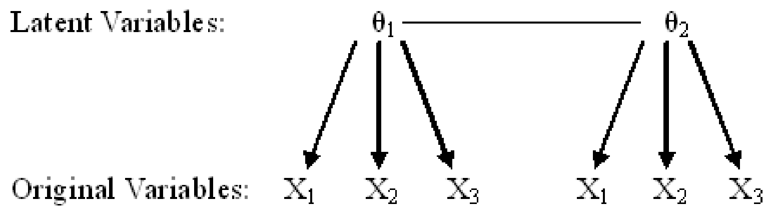
2.1.5 Latent Class Analysis (LCA)

The methods we discuss above either examine the customers aggregately or classify them into groups by some observed variables. In marketing practice where customer preference and utility are diverse, one model and one set of parameters do not hold true for all people (Rossi, Allenby and McCulloch, 2006). Suppose there are two customers who respond to the promotion only and one independent variable, for example, Profit. One customer has a value of +10 while the other has a value of -10. An aggregate estimation takes an average of the values of the independent variable and in this example, the variable Profit is considered have no effect since the average value is zero. From this example we can see that aggregate estimation diminish the influence of independent variables and customers with heterogeneity should be analyzed separately. Variables such as customer utility and demographics are often applied in the customer segmentation (Rossi and Allenby, 2003).

However, using observed variables may fail to thoroughly reflect the customer heterogeneity, and no single method can determine the appropriate variable for segmentation in direct marketing customer response prediction (Jedidi et al, 1997), It is necessary to identify some underlying structures to represent the communalities of the original variables. Since the underlying structures can not be observed directly, they are called latent structures and variables in these structures are called latent variables. The Latent Class Analysis (LCA) refers to the statistical process of finding out the latent classes or latent structures from multivariate categorical data (Heinen, 1996).

The main assumption of Latent Class Analysis is the “Local Independence”. That is, the observed variables are mutually independent given the latent variables. Before estimating the parameters, we have to specify the number of segments. Then observations are assigned to the latent classes according to the latent structure and different models are built for each latent class. Figure 4 illustrates a single structure of Latent Class Analysis.

Figure 4. Latent Class Analysis



In Figure 4, θ_1 and θ_2 represent the latent variables that define a 2-class latent group membership, in which the observations are homogenous on certain criteria. The arrows between latent variables and original variables indicate that the latent variables have direct influence on original variables. Coefficients and parameter values are the same within each latent class and differ from other latent classes.

Latent Class Analysis can discover the unobserved heterogeneity among customers in the direct marketing databases. Using latent variable formulation is not only flexible, but also able to generate any form of group membership (Rossi, Allenby and McCulloch, 2006). Researchers have shown the advantages of Latent Class Analysis over aggregate estimation in marketing issues for its ability of accounting for heterogeneity, as well as group size and probability (Jain, Bass and Chen, 1990). Previous studies also indicate the suitability of Latent Class Analysis in direct marketing applications (Wedel et. al., 1993).

However, there is no standard that how many latent classes should be included and different group settings lead to results which may significantly different from each other (Andrews and

Currim, 2003). Besides, the main assumption of "Local Independence" of Latent Class Analysis is often violated with real world data (Zhang, 2004), which also suggests a limitation of Latent Class Analysis application.

2.1.6 Artificial Neural Networks (ANNs)

Besides the techniques we mentioned above, data mining techniques such as Artificial Neural Networks (ANNs) have also been conducted in direct marketing. The ANNs are modeled to mimic and simulate the function of human brain, and are commonly used in direct marketing and related fields. ANNs consist of many non-linear computational elements called nodes, and different nodes arrange into different layers: the input layer, the hidden layer and the output layer. For example, in a 3-layer ANNs, the values of independent variables are inputted as the input nodes at first. Then, the input nodes are multiplied by the weights of the interconnections and these weighted inputs are algebraically added. After that, a nonlinear function, called a transfer or squashing function, is conducted and gives the results in the output nodes. This nonlinear function can be written in the following way:

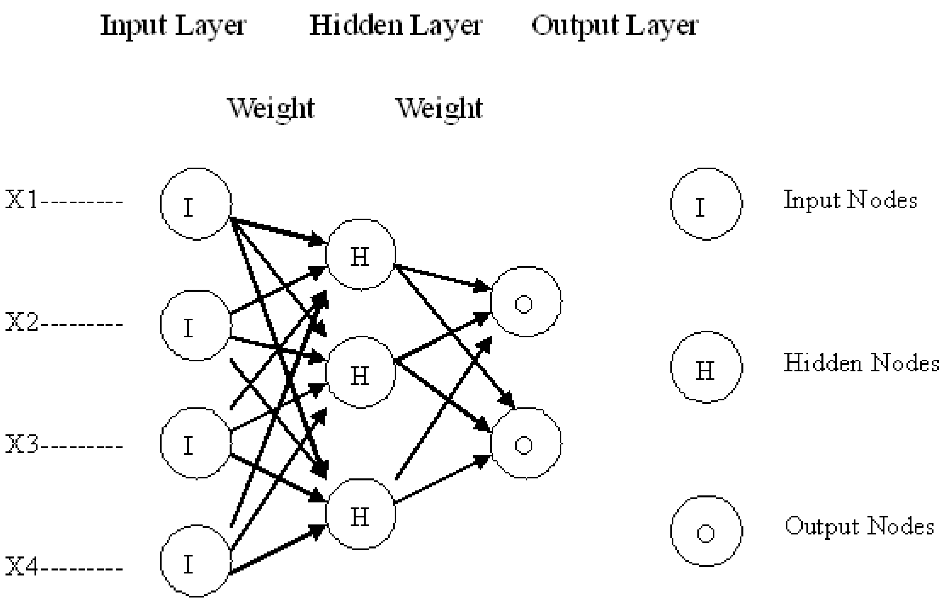
$$Y = \psi\left(\sum w_i X_i + \theta\right). \tag{2.1.3}$$

Here Y is the dependent variable, ψ is the transfer or squashing function, w_i is the weights or interconnection strengths, X_i is the independent variable and θ is the bias, which is similar to the intercept term in the Logistic Regression.

However, according to Shepard (1999), it is impossible to figure out the exact formula of this nonlinear function every time. If the training process is supervised, the function will find a best way to match the independent variables with the known dependent variable by minimizing the total error, which is the sum of squares of expected output values minus observed output values. In contrast, the unsupervised training process just uses the input values, and discovers the patterns without any external information. The whole ANNs are in fact multi-input nonlinear models with weighted interconnections between input and output layers which are shown in Figure 5 (Venugopal and Baets, 1994).

The ANNs do not make any assumption of the data distribution, and they can build their own model by discovering the underlying pattern of the independent variables. Thus ANNs are expected to perform better in classification task (Venugopal and Baets, 1994). Also they are able to properly handle data with missing values which is very difficult for regression, and they are good at operating nonlinear data and can identify the interactions between variables. These features make ANNs especially useful for the direct marketing database, which is often poorly structured (Cui and Wong, 2004).

Figure 5. Artificial Neural Networks



However, there are some problems when applying ANNs to direct marketing. First is the stochastic problem. Suppose that two observations with exactly the same attributes, and they will definitely be considered to have the same output. However, sometimes, this is not true in reality. These conflicts confuse the networks by sending contradictory information between the nodes and may cause the networks fail to converge to a stable set of weights (Zahavi and Levin, 1997).

Another problem is due to the large number of predictor variables in the direct marketing databases. Most of the predictor variables are based on the purchase history, redundancy occurs when some variables are highly correlated. It is also very difficult to decide which variables to use in any particular case (Zahavi and Levin, 1997), and ANNs also have the potential of model over fit (Cui and Wong, 2004).

2.1.7 Summary

We have discussed some of the frequently used classical direct marketing techniques. Some of these techniques are problematic in some perspectives. First is that these techniques fail to consider the uncertainty and assume the same parameter effect for the whole population. So they may build "good" models for the training data, but the models do not have the same predictive accuracy when applied to new test data. Second is that some the techniques neglect the customer heterogeneity. An aggregated estimation sometimes reduces the effect of independent variables by taking average values, and in most cases, separate models are required for different groups of customers. Third is the computational problem caused by the conflict of assumptions or disability of techniques themselves in analyzing large, noisy, and poorly constructed direct marketing data.

In this thesis, we apply a Bayesian approach in direct marketing customer response prediction and demonstrate how the Bayesian methods can overcome these problems.

2.2 Bayesian Methods in Direct Marketing

With the development of computation techniques and software, Bayesian methods have become widely used in direct marketing issues. Examples include the analysis of purchase time (Allenby, Leone, and Jen, 1999) and purchase frequency (Jen, Chou and Allenby, 2003), prediction of new product promotion (Neelamegham and Chintagunta, 1999), and choice model (Wedel et. al, 1999).

2.2.1 Direct Marketing Response Prediction

Although direct marketing problems are intensively reviewed in the marketing literature, a few articles are concerned with the response prediction (Steenburgh, Ainslie and Engebretson 2003). Bult and Wansbeek (1995) discuss the target customer selection by using a profit maximization approach. Steenburgh, Ainslie and Engebretson (2003) use the university admission campaign as a simulation of direct marketing issue and apply the Hierarchical Bayes method into the analysis with the incorporation of zip code information.

In Steenburgh and his colleagues' article (2003), they only compare the result of Hierarchical Bayes with that of the null model. More advanced model needs to be included in such comparisons. Moreover, they do not mention the variable selection process in which they choose four independent variables out of more than 200 demographic variables. The variable selection process itself is an important issue in direct marketing, and the results may become inconsistent when other sets of variables are put into the modeling process. Furthermore, the usage of student sample may seem inadequate in prediction since these students who are applying for admission have so many characteristics in common even they are from places of different zip codes. The admission process is a one-time event for the same student while direct marketing selling enrolls both historical and future promotions of individual customer. These two scenarios are quite different, and the conclusion drawn from the student sample has less applicability.

In this paper we apply the Bayesian methods in the direct marketing customer response prediction with an empirical direct marketing data. The results and conclusion drawn are considered more applicable.

2.2.2 Suitability of Bayesian Methods in Direct Marketing

The Bayesian methods are attractive in direct marketing problems for the following reasons. First is the variables selection process. Since different sets of sample data and different variable selection criteria or methods may lead to inconsistent combinations of selected variables, it is quite difficult to find a certain groups of predictor variables that have the same predictive accuracy when apply to datasets with new observations. The Bayesian methods can overcome this problem by incorporating the prior information, which is derived from the accumulated direct marketing data, and assume a random distribution of coefficient vectors, such as demographics, purchase histories, and behaviors among the customer population. Then the Bayesian methods make the inference of the response probability conditional on such distribution

(Rossi and Allenby, 2003). In other words, we can examine the historical data and discover a pattern that, which variables are more important in prediction than others. The Bayesian methods will combine the prior knowledge and information acquired from the data together into the inferences, and the setting of prior values is an indication of the researcher's belief of uncertainty. Such prior information is a unique feature of the Bayesian approach and is especially useful when there is little data based information or the ratio of such information to the parameters is low (Rossi and Allenby, 2003). The reason for this is that, the Bayesian methods make inferences based on the product of prior information and likelihood function, in which the latter can be obtained from the data. In both cases above where less information is available for likelihood estimation, the prior information has more influence on posterior estimation. If some variables are considered to be more meaningful, they are assigned a distribution with small variance, while a distribution with relatively large variance for variables that are not so important. The Bayesian methods take advantage of the prior information and the estimation process is not completely data oriented, thus the inference becomes more reasonable and the over fit of model can be reduced.

Second is the customer heterogeneity. Since customers are different from each other, we should analyze them differently and develop specific strategies for specific groups of customers. However, because of the information deficiency, researchers sometimes are unable to classify customers into groups, and they can only measure overall customers in an average way. Other methods divide customers into groups by criteria such as Chi-Square, and build models separately. Hierarchical Bayes, which also takes advantage of prior knowledge, separates customers into different groups by some factors and analyzes them differently. The Hierarchical

Bayes differs from other methods in which although different models are constructed for different groups, these models are not independent from each other, and they will modify themselves conditional on the information of other models (Allenby, Bakken and Rossi, 2004).

Third is the ability of analyzing large and noisy data. The Bayesian methods are free from the assumptions of the distributions and the types of the data, and are capable of analyzing poorly constructed direct marketing data (Cui, Wong and Lui, 2006). The prior information can also provide better solution to high dimension data inference and help to identify the most influential variables (Genkins, Lewis and Madigan, 2005).

3 Bayesian Methods

During the past several years, there has been a significant increase of Bayesian methods applications in the marketing context, from the new product introduction to pricing (Rossi, Allenby and McCulloch, 2006). Although the Bayesian methods have been used in areas other than marketing at first, the recent development of computational methods makes Bayesian methods more applicable. Besides, the improvement of marketing data accessibility also provides researchers and practitioners with adequate information for Bayesian inferences.

We do not assume the universal applicability of Bayesian methods in marketing issues. According to the No Free Lunch Theorem, no single method is definitely better than others in any given situation (Ho and Pepyne, 2002). We propose the suitability of Bayesian methods in direct marketing context because first, the large volume of customer transaction record can help to assess the prior information which the Bayesian methods need to make future prediction. Second, with the diversification of customer preference and utility, a separate analysis of customers is required, especially in direct marketing where an individual level of decision is made. Bayesian methods is capable of examine the customer heterogeneity and build less aggregate models through methods such as Hierarchical Bayes.

3.1 Bayesian Statistics

All Bayesian methods are theoretically based on the Bayes' Theorem, which was proposed by Thomas Bayes in the 1760s. Although Bayes' Theorem has long been conceptually appealing, it is not widely applied due to the computational constraint. However, such a constraint is solved with the emergence of simulation methods such as Markov Chain Monte Carlo (MCMC). In this section, we mainly discuss the Bayes' Theorm, the Prior Information which makes Bayesian methods attractive, and the simulation method MCMC.

3.1.1 Bayes' Theorem

The Bayes' Theorem can be written in the following form:

$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{P(B)} = \frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{\infty} P(B \mid A_i)P(A_i)} \tag{3.1.1}$$

In formula 3.1.1, $P(A_j)$ is the prior probability or marginal probability of event A_j, $P(A_j \mid B)$ is the conditional probability of event A_j given event B and it is also called the posterior probability, $P(B \mid A_j)$ is the conditional probability of event B given event A_j, and $P(B)$ is the prior or marginal probability of event B and acts as a normalizing constant. Thus, the Bayes' Theorem can be considered in another way, that is, the posterior probability equals to the product of likelihood estimation and prior probability, divided by a normalizing constant, shown in formula 3.1.2.

Posterior Probability = Likelihood × Prior / Normalizing Constant (3.1.2)

The Bayes' Theorem assumes that, given a set of hypotheses, each one has a certain probability of being correct or being an event. Receiving more information will change the probabilities from a learner's point of view. For instance, an observation may be contradictory to a hypothesis, or strengthen the belief when we get a deeper understanding of it. The aim of this setting is to find a hypothesis with the highest probability of being correct or being an event, given a specific set of data or information. This assumption is different from the rule of the classical statistical methods which calculate the probability by using the number of occurrences of specific events divided by the total number of total events. The Bayes' Theorem considers that the probability has nothing to do with the event that has happened, and it tries to identify some other "related" information and improve the estimation of the probability.

3.1.2 Prior Information

As we mentioned before, the Bayesian approach to make prediction or estimation is based on some existent knowledge or prior information. Such a statement of prior information is a major difference between Bayesian methods and some other methods.

There are several types of prior, and we can classify them into two major categories: the subjective prior, and the objective prior. By subjective prior, we mean that the prior information is acquired from some experts' opinions or experiences. Such prior information may sometimes be biased and fail to reflect the underlying "truth". However, since the information we have is far from enough, the subjective prior information is undoubtedly valuable in the estimation process. Besides, in some other occasions, the prior information can be achieved from data, especially historical dataset (Cui, Wong and Lui, 2006). The prior information can be used to better infer the pattern of some specific behaviors, and such inference can provide us more knowledge and help us make a more accurate prediction. That is why we consider the Bayesian methods are suitable for direct marketing data analysis since the direct marketing database often contains some historical purchase records, which can help us to understand more about the customers' behavior.

The historical purchase records are also considered more meaningful indicators than other information such as the customers' demographics (Rossi, McCulloch and Allenby, 1996). Our empirical study result also indicates that transaction history more important in modeling building than credit information and demographics.

The settings of the forms of prior and hyperparameter, such as prior variance, are very important in Bayesian applications because they can have significant influences on individual level estimation (Rossi and Allenby, 2003). Researchers have investigated various distributions, such as the Poisson distribution (Neelamegham, and Chintagunta, 1999), generalized gamma distribution (Allenby, Leone, and Jen, 1999), and Laplace distribution (Genkin, Lewis and Madigan, 2005). Rossi and Allenby (2003) also argue that a specification of normal distribution for the prior is appropriate because it can diminish the influences of outliers. The hyperparameter setting depends on the degree of prior beliefs. Normally speaking, the small variance often indicates that the independent variables have relatively more consistent effects over the population than other variables which have larger prior variances. However, because of the data insufficiency and the cost of assessing prior knowledge, sometimes it is difficult to acquire the actual values for the hyperparameters. Genkin et al. (2005) propose a geometric setting for the hyperparameters. That is, we assign a set of values for the hyperparameters and compute the posterior probability conditional on each set of values. The set which can maximize the posterior probability is chosen as the hyperparameter values.

Despite the difficulty and cost, the prior information is definitely important in data analysis, and that is why Bayesian methods generally have a better performance than other methods in cases of high data uncertainty (Rossi and Allenby, 2003).

3.1.3 Markov Chain Monte Carlo (MCMC)

Because of the computational difficulty, the posterior calculation of Bayesian methods was once seemed impossible. Simulation methods such as the Markov Chain Monte Carlo (MCMC) have been developed to solve this problem.

There are a lot of comprehensive reviews of MCMC simulation method (Robert, and Casella, 2004) and we do not include the details here. Basically speaking, given the prior information and likelihood, the Bayesian methods explore the posterior distribution. In many cases, the posterior distribution is not normal and the problem is to estimate the distribution through a simulation method (Rossi, Allenby and McCulloch, 2006). However, the estimation is infeasible because of the computational constraint and MCMC solves this problem using a Markov chain. It makes simulations or iterations using equivalent distribution. Then it makes draws from a set of random variables and revises the values of the parameters through the iteration processes until convergence conditions are met. The MCMC simulation method is now considered an appropriate solution for Bayesian computation (Rossi and Allenby, 2003).

Two MCMC methods are widely used, one is the Gibbs sampler, and the other is the Metropolis methods. Gibbs sampler has been applied in binary probit, mixture of normals and hierarchical linear models, and Metropolis methods are used for multinomial logit model (Rossi, Allenby and McCulloch, 2006). In direct marketing customer response prediction where the dependent variable is a binary one, the Bayesian methods we apply in this study use the Gibbs sampler.

3.2 Bayesian Methods in Direct Marketing Response Prediction

Two Bayesian methods are conducted in this article. We first apply the Bayesian Binary Regression to model customers as a whole group, and several Hierarchical Bayes models using different clustering variables are used to account for the customer heterogeneity.

3.2.1 Bayesian Binary Regression (BBR)

Regression analysis is widely used in marketing research for inferring the relationship between independent variables and dependent variables. Since in direct marketing database, the dependent variable is a binary variable with values of 0 and 1 representing “nonrespondents” and “respondents” respectively, thus the binary logistic regression is often considered a suitable method for analysis.

BBR is in fact a Logistic Regression with the incorporation of prior information, which applies the prior information in the estimation and modification of parameter values throughout the iteration process. The whole computation process of BBR is similar to that of Logistic Regression, but the underlying algorithm is not the same. Logistic Regression uses Maximum Likelihood Estimation (MLE) to calculate the parameter coefficients while BBR uses the prior information and posterior mode of the training data to find the Maximum Posterior (MAP) in test data.

Specifically, in direct marketing customer response prediction, the dependent variable is the binary variable of the group membership of customers: usually with a value of “1” for respondent and “0” for nonrespondent. Independent variables such as purchase history, demographics, and credit information are denoted by $x_1, x_2, x_3 \ldots x_n$. According to the Bayes Theorem, the posterior probability estimation equals to:

Pr (R = 1| H, D, C) = Pr (R = 1) * Pr (P, D, C | R= 1) / Pr (P, D, C) (3.1.3)

Here R refers to Response with the value of 1 for respondent and 0 for nonrespondent. H means Purchase History data, D means Demographic information, and C means Credit information.

Since the numerator can be considered a normalized constant and has no significant influence on the posterior inference, we may be just concerned with the denominator part. In most cases, the independent variables are assumed to be independent with each other (which may not be true in reality), and then the denominator part can be written as:

Pr (R= 1) * Pr (P, D, C | R = 1) = Pr (R = 1, P, D, C) = Pr (R = 1) * Pr (P | R= 1) * Pr (R= 1) * Pr (D | R = 1) * Pr (R = 1) * Pr (C | R = 1) (3.1.4)

In the other way, the whole equation can be simplified as:

Pr (Y=1| $x_1, x_2, x_3 \ldots x_n$) = Pr (Y=1) * Pr (x_1 |Y=1) * Pr (x_2 |Y=1) * Pr (x_3 |Y=1)… *Pr (x_n |Y=1) or Pr (Y=1) * $\prod_{i=1}^{n} \Pr(x_i | Y = 1)$ (3.1.5)

Among all the independent variables, the purchase history is considered the most important indicator (Rossi, McCulloch and Allenby, 1996), especially the RFM variables (Gonul and Hofstede, 2006). A simple test result shows that, when we include all independent variables in the modeling process, the purchase history data, such as the order quantities of the same promotion last year, the elapsed time since last purchase, and the money spent have the most significant effects. However, realizing that different purchase pattern is related with the characteristic of individual customer, we may also add some other demographic and credit information in the model to make the result more meaningful.

When only those “important” variables are included in the model, the effects are significant and the coefficients are significantly different from zero, thus we assume a normal distribution of priors among the customer population, which is considered an appropriate approach (Rossi, and Allenby, 2003).

BBR is able to incorporate the prior information and account for the uncertainty in direct marketing customer response prediction. Since it is unlikely that the independent variables have the same effects over the population in reality, the results obtained from BBR seem to be better than those obtained from other aggregate models such as logistic regression. For variables that have more predictive accuracy, relatively smaller prior variances are assigned. While variables have less predictive accuracy, relatively larger prior variances are assigned.

3.2.2 Hierarchical Bayes (HB)

Since customers are heterogeneous, a more appropriate way in the modeling process is to build separate models for groups of customers. The Hierarchical Bayes is good at accounting for the heterogeneity and is frequently applied in marketing issues (Rossi and Allenby, 2003).

The HB model is called “hierarchical” because it has two stages and it calculates the posterior probability through the production of the unit level likelihood and two stages of priors. The first stage prior is the overall prior information of the data, and the second stage prior is the individual prior information given the group membership. How to separate customers into different hierarchies depends on the specific problems themselves. When we are interested in the customer decisions or behaviors, “utility” is often considered a very important predictor since customers’ preferences are different towards the products (Allenby, Bakken and Rossi, 2004). In some other cases, variables such as demographics, purchase records, or credit information can also be used as the criteria to distinguish different groups of observation or customers. Rossi et al. (2006) also suggest the application of latent variables. They list several advantages of using latent classification. First, the latent classification can generate any kind of discrete models. Second, the latent classification makes it easier for MCMC algorithms using data augmentation. Third, the latent classification can have a random utility interpretation which is related to the latent variables other than utility maximization.

The Hierarchical Bayes model estimation starts at an aggregate level: assigning each individual customer with some basic parameter values and revising these values throughout the iteration process. Although different models are constructed for different groups of customers, each individual customer’s parameter estimation is not independent, and the parameter values are adjusted according to other customers’ information. Because of this, the parameter estimation at the individual level is more stable and better in reflecting the specific pattern of each customer or observation.

Hierarchical Bayes assumes the uncertainty using prior information and allows differences within each group. Latent Class Analysis assumes that the independent variables have same effect on each latent group, and the heterogeneity accounted by Latent Class Analysis is a discrete one. Hierarchical Bayes assumes a continuous heterogeneity both between and within groups, and the estimation is more individualized. While these two segmentation techniques have been widely discussed, researchers have not reached an agreement on which one is better. Since in direct marketing customer response prediction, the decision is made in an individual level (Wedel et al, 1999), we believe the Hierarchical Bayes is more appropriate.

Now the problem is which variables should be used for Hierarchical Bayes model classification because different variables chosen may lead to estimation results that significantly different from each other. Researchers propose several classification criteria but which criterion is better is not well understood. The selection of classification methods depends on the specific characteristic of the problems (Andrews and Currim, 2003). Direct marketing database provides with adequate demographic and geographic information for classification, while the latent variable is also an appropriate choice. In this study, we conduct Hierarchical Bayes models using both observed variables and latent variables and compare the model performances.

Figure 6 shows the different effects of independent variables, and Table 1 illustrates some differences among these techniques.

Figure 6. Different Influences of Independent Variables

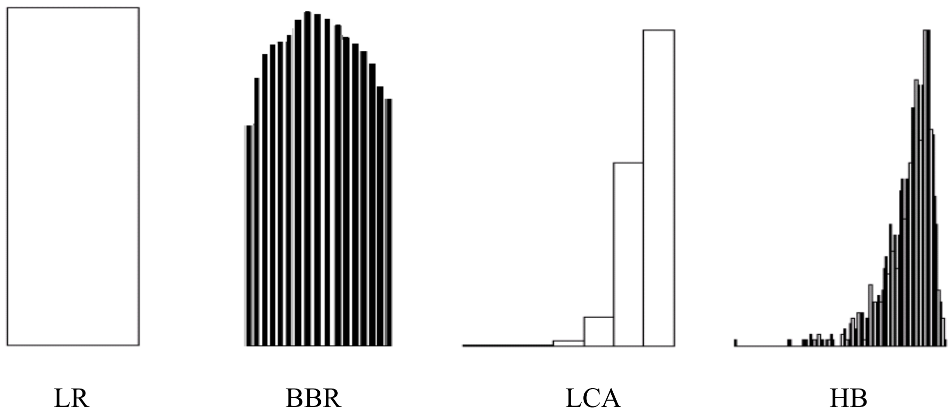


Table 1. Different Features of LR, BBR, LCA, and HB.

Features	LR	BBR	LCA	HB
Uncertainty	No	Yes	No	Yes
Between Group Heterogeneity	No	No	Yes	Yes
Within Group Heterogeneity	No	Yes	No	Yes

From Figure 6 and Table 1 we notice that, for logistic regression, the independent variables have the same effects over the population. For Bayesian Binary Regression, the effects of independent variables form a certain kind of distribution over the population. In Latent Class Analysis where separate models are built, the independent variables have different influences among the latent classes but the same effects within each class. For Hierarchical Bayes, the effects of independent variables vary not only among groups, but also within each group.

4 Data Analysis

We apply four different techniques in our study using the same training and test datasets, which are randomly chosen from an empirical direct marketing data, and compare their model predictabilities. Techniques applied include Logistic Regression (LR), Bayesian Binary Regression (BBR), Latent Class Analysis (LCA) and Hierarchical Bayes (HB).

4.1 Empirical Direct Marketing Data

The dataset we use in this article obtained from an American catalogue company which sells multiple products, such as gifts, apparel, electronics, and houseware. The company sends regular mailings to its list of customers and keeps a longitudinal record of every customer. Besides transaction history, the dataset also combines the zip code level credit information and customer demographics. The dataset is provided for the purpose of academic research. Examples using this dataset include Cui et al. (2006).

4.1.1 Data Description

The whole dataset contains 103,713 observations and 307 independent variables. The total response rate is 5.34%, that is, among those 103,713 customers or households who have received the promotional material, only 5,539 of which who finally purchase the products. The dependent variable is a binary variable with “1” for respondent and “0” for nonrespondent. The independent variables include demographic variables such as gender, education level, income, and social status. Credit information such as bank credit limit and bank balance, and purchase history records such as total number of purchases per product, recency, money spent, and responses toward promotions in the past, are also included. For the purpose of privacy protection, the product categories included in the promotion material are not mentioned, and customer credit and demographic information is averaged based on the zip code level. Detail information of the independent variables can be found in Appendix 1.

We use the methods mentioned above to find out whether the people or households who have received the mail will make a purchase or not. Those customers who have the highest probabilities of making a purchase will be considered as the target customers.

4.1.2 Train and Test Data

In order to reduce the probability of over fitting of any single model, 10 holdout experiments are conducted using 10 datasets. These 10 datasets are similar in size and are randomly chosen from the original dataset, namely 01A, 01B, 02A, 02B…05A, and 05B. All these datasets act as both train and test data. For example, we use 01A for model building and 01B for validation in one experiment, and we use 01B for model building and 01A for validation in another. We use 10 holdout experiments rather than 10 fold cross validation because the latter is much more time consuming. All training and testing datasets are mutually exclusive. Detail information of the number of respondents and nonrespondents in the datasets can be found in Table 2.

Table 2. Train and Test Data: Original Ratio

Dataset	Number of Respondent	Number of Nonrespondent	Total
01A	281	4873	5154
01B	272	4945	5217
02A	288	4888	5176
02B	267	4928	5195
03A	265	4911	5176
03B	256	4939	5195
04A	285	4855	5140
04B	278	4953	5231
05A	287	4847	5134
05B	280	4957	5237

Table 3. Train Data: Balance Ratio

Dataset	Number of Respondent	Number of Nonrespondent	Total
01A’	281	281	562
01B’	272	272	544
02A’	288	288	576
02B’	267	267	534
03A’	265	265	530
03B’	256	256	512
04A’	285	285	570
04B’	278	278	556
05A’	287	287	574
05B’	280	280	560

Since the ratio of positive response is quite low, the nonrespondents may dominate the estimation process. In order to eliminate the effect of the high proportion of nonrespondents, we randomly delete some of the observations who respond negatively to the promotion in the train data to make a balance between respondents and nonrespondents. The new created balanced ratio training data are named as 01A’, 01B’, 02A’, 02B’…05A’, and 05B’, and another 10 experiments are conducted using the balanced ratio training data. Detail information can be found in Table 3.

4.2 Parameter Setting

For Bayesian Binary Regression, the most important parameter setting includes the form of prior distribution and the hyperparameter values. For Hierarchical Bayes, besides these two settings, customer group membership needs to be assigned before the estimation.

4.2.1 Prior Setting

For both Bayesian Binary Regression and Hierarchical Bayes, we set the prior form the normal distribution which is proposed by Rossi and Allenby (2003). A geometric setting of values of prior variance proposed by Genkin et al. (2005) is applied here. The prior variance values range from 0.001 to 1000, indicating the belief of the effects of different independent variables. A 10 fold cross validation is conducted to choose the best set of prior variances that maximize the posterior probability.

We admit that the geometric setting of prior variance does not include all possible values. However, the independent variables contained in the estimation are considered meaningful in prediction and their effects on dependent variable do not vary significantly over the population. So the actual values of the variance do not significantly differ from proposed values and the estimation results are stable.

Table 4 and Table 5 illustrate examples of the selection of prior variances based on the cross validation results of the BBR and HB respectively. In general, small variances of the prior results in better log likelihood statistics.

Table 4. Prior Variance Selection of BBR

	Variance	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
LL	Fold 1	-32.52	-33.27	-33.74	-34.13	-34.45	-34.74	-35.00	-35.24	-35.45	-35.65
	Fold 2	-34.97	-35.75	-36.48	-37.05	-37.60	-38.11	-38.57	-38.99	-39.40	-39.78
	Fold 3	-34.85	-35.86	-36.84	-37.59	-38.24	-38.80	-39.31	-39.77	-40.19	-40.58
	Fold 4	-36.31	-37.23	-38.25	-39.06	-39.75	-40.36	-40.89	-41.37	-41.80	-42.19
	Fold 5	-28.30	-27.92	-27.83	-27.86	-27.93	-28.01	-28.11	-28.20	-28.29	-28.38
	Fold 6	-27.07	-27.42	-27.90	-28.37	-28.78	-29.15	-29.49	-29.79	-30.07	-30.32
	Fold 7	-35.03	-35.93	-36.47	-36.87	-37.18	-37.43	-37.64	-37.82	-37.98	-38.11
	Fold 8	-32.04	-32.75	-33.30	-33.68	-33.99	-34.26	-34.48	-34.68	-34.85	-35.01
	Fold 9	-33.92	-34.04	-34.54	-34.93	-35.28	-35.62	-35.94	-36.25	-35.53	-36.80
	Fold 10	-47.36	-49.77	-51.67	-53.21	-54.42	-55.45	-56.32	-57.08	-57.76	-58.38
	Mean	-34.24	-34.99	-35.70	-36.28	-36.76	-37.19	-37.58	-37.92	-38.13	-38.52

LL=Log Likelihood

Table 5. Prior Variance Selection of HB

Prior Variance		Mean Log Likelihood of 10 fold Cross Validation	Standard Error
Level 1	Level 2		
0.001	0.001	-27.24	2.12
0.001	0.002	-26.78	2.21
0.001	0.003	-26.44	2.26
0.002	0.001	-26.85	2.19
0.002	0.002	-26.47	2.25
0.002	0.003	-26.19	2.30
0.003	0.001	-26.55	2.25
0.003	0.002	-26.21	2.29
0.003	0.003	-25.97	2.33

When geometric setting of prior variance is applied in BBR training, the training data are separated into 10 sub data, in which 9 sub data are contained in the modeling using all prior variances provided while the remaining one is used for validation. The process repeats for 9 times and all posterior values are compared. The prior variance that has the largest posterior value over the 10 fold cross validation is chosen in the modeling process. Suppose we are choosing from a set of variances 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. From Table 4, we notice that in this particular example, the posterior value of model using prior variance 0.1 is the largest. Thus we set the prior variance equals to 0.1 in the estimation.

The selection of prior variance in HB is similar with that in BBR but there are some important differences. HB has two levels of priors and the posterior estimation is based on the combination of both their variances. From Table 5, we notice that, when both level 1 and level 2 prior variances equal to 0.003, the average posterior has the largest value over the 10 validation folds. Thus we choose 0.003 and 0.003 from the setting of two levels of prior variances (level 1: 0.001, 0.002, and 0.003; level 2: 0.001, 0.002, and 0.003) in this example.

4.2.2 Hierarchy Setting

Evans et al. (2000) investigate the gender effect on direct marketing response and discover the different response patterns between female and male customers. Customers can also be classified according to their geographic location, because people tend live together with others who are similar with themselves (Steenburgh, Ainslie and Engebretson, 2003). Moreover, economic status is also considered related with the buying behavior which may affect the responses. So, three Hierarchical Bayes models are built using the demographic variables Gender, State, and Wealth. Gender has values of 1, 2, and 3, which indicate Female, Male, and Company respectively. State ranges from 1 to 51, indicating 51 different states of the U.S. Wealth ranges from 0 to 9, from lowest to highest, indicating the relative wealth level of each observation.

Another two Hierarchical Bayes models are built using latent variables which are defined by

Latent Class Analysis. The numbers of latent classes are 2 and 3 respectively. We have discussed the advantages of using latent variables in Hierarchical Bayes and researchers also prove the improvement of model accuracy through the incorporation of latent variables in Hierarchical Bayes (Langseth and Nielsen, 2006). In our study, we do not use the Hierarchical Bayes to generate the latent variables which is currently impractical. Instead, we “borrow” the latent class membership from the result of Latent Class Analysis.

We have to address that the main purpose of this study is the customer response prediction, not the comparison of abilities of different models choosing latent variables. Besides, even using the same dataset, the Latent Class Analysis does not make exactly the same classification in every experiment. Also, how many latent classes should be defined is not the concern here. We simply test the 2-class and 3-class latent variables models for practical reasons, and we understand that these two choices may not be the optimal one. We do not want the difference in model performance is caused by the different grouping of customers. Instead, only when customers are assigned in the same latent classes can we discover the difference between the Hierarchical Bayes using latent variables and the Latent Class Analysis.

4.3 Response Prediction

After the modeling process, the modeling result itself has limited usefulness because we can not guarantee the model can be applied to new observations. We have to use the test data file to make validation.

4.3.1 Classification Error

Classification Error is a standard to evaluate the performance of classification models. The error rate is calculated using the following formula. Normally speaking, the less the classification error rate, the better the model predictive accuracy.

Total Error Rate = No. of Wrong Labeling / Total No. of Observation (4.1.1)

However, in direct marketing response prediction, the Classification Error can not truly reflect the model performance. Here we apply the Confusing Matrix in Table 6 to explain it.

Table 6. Confusing Matrix

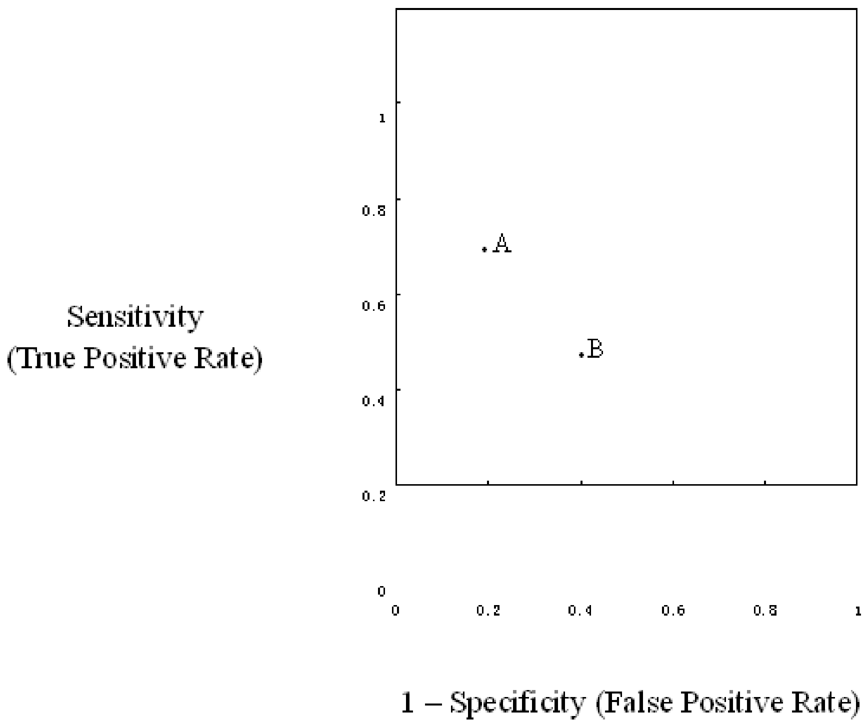
		Correct Label	
		1	0
Prediction	1	True Positive	False Positive
	0	False Negative	True Negative

According to the Confusing Matrix, if a customer is predicted to be a respondent/nonrespondent and in fact he or she purchases/does not purchase the product, then such prediction is considered true positive/true negative". On the contrary, if a customer is predicted to be a nonrespondent/respondent, but in fact he or she purchases/does not purchase the product, then such prediction is considered false negative/false positive. Obviously, the cost of false negatives is much greater than that of false positives (Cui, Wong and Lui, 2006). That is why nowadays the simple error rate is no longer the most appropriate method for assessing the model performance.

4.3.2 Receiver Operating Characteristic (ROC)

Because the simple classification error does not truly reflect the model performance, another method called Receiver Operating Characteristic (ROC) has been applied to select the classification methods based on their performances. The Receiver Operating Characteristic or simply ROC curve is a graphic representation of the trade off between the true positive rate and false positive rate (Fawcett, 2006). While first applied in medicine, radiology, and psychology, the ROC curve has been introduced in other areas such as machine learning and data mining recently. Figure 7 illustrates the ROC graph or ROC space.

Figure 7. ROC Space



In Figure 7, the vertical axis refers to the sensitivity or true positive rate of the model and the horizontal axis refers to the 1 – specificity or false positive rate of the model. The true positive rate and false positive rate can be calculated from the following formulas.

True Positive Rate = Positives Correctly Classified / Total Positives (4.1.2)

False Positive Rate = False Negatives / (False Positives + True Negatives) (4.1.3)

The point (0, 0) indicates the method that makes no false positive errors but also has no true positives. The point (1, 1) indicates the method which classifies all observations as positives. The point (0, 1) indicates the perfect classification where the true positive rate is 100% and no false positive occurs. One point in ROC space is better than another if it is on the upper left area (point A and point B in Figure 7), which indicates higher true positive rate and lower false positive rate (Fawcett, 2006).

Discrete classifiers, such as decision tree, result in single point in the ROC space when an observation is given. Other methods produce probability values to represent the degree to which group the observation belongs to. In our case, all methods used belong to the second type. Different cut off values can be used to make a trade off between the true positive rate and false positive rate. When infinite numbers of cut off values are used, the result in ROC space will be a curve. Researchers have discussed the interpretation of the ROC curve. Normally speaking, the larger the area under the ROC curve, the better the predictive accuracy (Fawcett, 2006).

4.3.3 Cumulative Lift Value

Besides ROC curve, the cumulative lift value can also reflect the model performance. In order to calculate the cumulative lift value, we first rank the observations in the test according to their predicted probability and evenly divide them into 10 deciles. As we mentioned before, customers with higher predicted probability are considered target customers. We calculate the number of true positives in each decile and compare it with the total number of true positives. Table 7 illustrates an example of cumulative lift value result.

Table 7. Cumulative Lifts as Results

Decile	Records	Prob (1)	% (1)	Cum. % (1)	# (1)	Cum. # (1)	%Tot (1)	Cum.% Tot (1)	Lift	Cum. Lift
0	521	0.9617	23.42	23.42	122	122	44.85	44.85	449.132	449.132
1	521	0.7590	12.09	17.75	63	185	23.16	68.01	231.929	340.530
2	521	0.5130	6.14	13.88	32	217	11.76	79.78	117.805	266.289
3	521	0.3472	3.07	11.18	16	233	5.88	85.66	58.903	214.442
4	521	0.2371	2.69	9.48	14	247	5.15	90.81	51.540	181.862
5	521	0.1594	0.77	8.03	4	251	1.47	92.28	14.726	154.006
6	521	0.1037	1.15	7.05	6	257	2.21	94.49	22.089	135.160
7	521	0.0607	0.58	6.24	3	260	1.10	95.59	11.044	119.646
8	521	0.0294	0.96	5.65	5	265	1.84	97.43	18.407	108.397
9	528	0.0053	1.33	5.21	7	272	2.57	100.00	25.428	100.000

In Table 7, Records indicate the number of observations in each decile. Prob (1) represents the average predicted probability of positive observations in each decile. % (1) means the percentage of positive observations over the total observations in each decile. Cum. % (1) indicates the cumulative percentage of positive observations over the total observations. # (1) represents the number of positive observations in each decile. Cum. # (1) means the cumulative number of positive observations. %Tot (1) indicates the percentage of positive observations over the total number of positive observations in each decile. Cum. %Tot (1) represents the cumulative the percentage of positive observations over the total number of positive observations. Lift means the 100 ratio of accuracy over the random model in each decile, and the Cum. Lift indicates the cumulative 100 ratio of the accuracy over the random model. For example, the number 449.132 means the model performs 4.49 times as well as the random model in the first decile. The higher the cumulative lift value in the upper deciles, the better the model predictive accuracy. Due to the budget constraint, normally only the first one or two deciles are considered (Cui, Wong and Lui, 2006).

4.4 Results

A variable selection is conducted using the Backward method first and Forward method then through SPSS 15.0. Besides, some independent variables that excluded from the variable selection process are also put into analysis. The reason for doing this is that, these variables are considered meaningful in response predicting according to some subject or object prior knowledge. Finally, 109 independent variables are included in the data analysis, and the detail information can be obtained in Appendix 2.

Four techniques, Logistic Regression, Bayesian Binary Regression, Latent Class Analysis, and Hierarchical Bayes, are applied using the same train and test data. The software we use in this study include SPSS 15.0 for Logistic Regression, Latent Gold 3.0 for Latent Class Analysis, and BBRtrain and BBRtest, which are encoded using C language, for Bayesian Binary Regression and Hierarchical Bayes analysis.

We first conduct 10 experiments using all techniques with the original ratio training and testing data. In order to check whether the ratio of respondents and nonrespondents has effect on the model performance, we further conduct another 10 experiments using those techniques with the balanced ratio training and original ratio testing data. The respondents in both original ratio training and balanced ratio training data are the same, while the nonrespondents of the balanced ratio training data are randomly chosen from the nonrespondents in the original ratio training data. Table 8 and Table 9 show the detail information of the experiments.

Table 8. Experiment Information of Training Data with the Original Ratio

Experiment	Original Ratio					
	Train Data	No. of 1	No. of 0	Test Data	No. of 1	No. of 0
1	01A	281	4873	01B	272	4945
2	01B	272	4945	01A	281	4873
3	02A	288	4888	02B	267	4928
4	02B	267	4928	02A	288	4888
5	03A	265	4911	03B	256	4939
6	03B	256	4939	03A	265	4911
7	04A	285	4855	04B	278	4953
8	04B	278	4953	04A	285	4855
9	05A	287	4847	05B	280	4957
10	05B	280	4957	05A	287	4847

Table 9. Experiment Information of Training Data with Balanced Ratio

Experiment	Original Ratio					
	Train Data	No. of 1	No. of 0	Test Data	No. of 1	No. of 0
11	01A’	281	281	01B	272	4945
12	01B’	272	272	01A	281	4873
13	02A’	288	288	02B	267	4928
14	02B’	267	267	02A	288	4888
15	03A’	265	265	03B	256	4939
16	03B’	256	256	03A	265	4911
17	04A’	285	285	04B	278	4953
18	04B’	278	278	04A	285	4855
19	05A’	287	287	05B	280	4957
20	05B’	280	280	05A	287	4847

4.4.1 Classification Error

Table 10 and Table 11 show the Log Likelihood of original ratio and balanced ratio training data respectively, and Tables 12 to 17 illustrate the Total Classification Error, Respondent Classification Error and Nonrespondent Classification Error of the models respectively. The term “Total Classification Error” means the sum of false positives and false negatives over the total number of observations. “Respondent Classification Error” and “Nonrespondent Classification Error” represent the false negative rate and false positive rate respectively.

Table 10. Log Likelihood of Training Data with the Original Ratio

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
1	-813.251	-817.978	-796.046	-478.500	-705.846	-402.112	-374.598	-639.609	-534.241
2	-753.095	-834.392	-770.476	-439.742	-661.658	-781.320	-219.132	-574.719	-418.991
3	-761.322	-768.858	-763.953	-459.114	-667.260	-342.503	-766.556	-597.336	-470.356
4	-772.970	-778.055	-765.449	-460.138	-685.620	-413.588	-778.406	-613.565	-464.952
5	-769.878	-807.178	-800.530	-510.094	-714.048	-342.799	-799.326	-600.534	-501.970
6	-737.572	-739.535	-720.256	-448.209	-655.289	-744.904	-149.898	-561.938	-446.944
7	-786.690	-797.444	-790.045	-476.411	-692.544	-371.300	-800.827	-629.580	-531.444
8	-799.538	-869.743	-799.601	-496.563	-728.930	-334.138	-250.355	-648.347	-490.885
9	-828.955	-834.783	-852.825	-510.766	-752.872	-859.763	-295.520	-647.418	-498.216
10	-766.087	-851.746	-774.289	-433.611	-665.144	-317.988	-315.337	-564.980	-465.234
Mean	-778.936	-809.971	-783.347	-471.315	-692.921	-491.042	-474.996	-607.803	-482.323
SD	28.158	40.040	34.185	27.891	32.286	213.831	274.433	33.211	36.240

In all result files of this thesis, LR means Logistic Regression, BBR means Bayesian Binary Regression, HBG means Hierarchical Bayes using Gender as clustering variable, HBS means Hierarchical Bayes using State as clustering variable, HBW means Hierarchical Bayes using Wealth as clustering variable, HBL2 means Hierarchical Bayes using 2-class latent variable for clustering, HBL3 means Hierarchical Bayes using 3-class latent variable for clustering, LCA2 means 2-class Latent Class Analysis and LCA3 means 3-class Latent Class Analysis. SD means the standard deviation.

Table 11. Log Likelihood of Training Data with Balanced Ratio

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
11	-236.050	-261.949	-232.261	-70.162	-159.498	-169.816	-163.871	-100.132	-94.238
12	-200.373	-243.707	-219.609	-64.071	-164.992	-170.013	-136.762	-88.411	-87.584
13	-199.182	-275.408	-200.103	-72.983	-178.307	-147.768	-136.651	-86.788	-81.862
14	-168.895	-223.625	-190.240	-58.146	-142.982	-124.246	-127.610	-72.323	-73.941
15	-209.214	-261.898	-220.180	-65.317	-154.036	-178.423	-165.147	-94.556	-91.137
16	-171.604	-230.322	-184.194	-68.666	-134.620	-147.198	-117.216	-66.604	-74.378
17	-208.581	-260.350	-228.023	-80.624	-161.632	-214.154	-145.018	-94.003	-86.438
18	-191.315	-235.124	-225.945	-75.615	-187.851	-161.179	-38.613	-77.237	-75.731
19	-227.776	-272.401	-248.039	-92.583	-150.453	-206.193	-136.427	-98.946	-93.343
20	-225.783	-258.404	-266.151	-89.557	-165.199	-197.397	-148.850	-105.073	-96.696
Mean	-203.877	-252.319	-221.475	-73.772	-159.957	-171.639	-131.617	-88.407	-85.535
SD	22.593	17.949	25.184	11.070	15.727	28.362	35.885	12.723	8.603

We list the model coefficients of BBR, LR, HBL2, HBL3, LCA2, and LCA3 in Appendix 3 and 4. These coefficients result from one experiment using original ratio training data and another experiment using half ratio training data. From the model coefficient results we notice that, different techniques make different estimations, not only the significance of the effect of independent variables, but also the direction of the influence in some cases when using the same data. For LCA, the coefficient estimations for each latent class are quite dissimilar, while for HB the coefficient estimations for each latent class are similar. This certifies that LCA makes different models for different classes, and the models built by HB share the information among themselves and modify the coefficient values through the estimation process.

Table 12. Total Classification Error: Training Data with the Original Ratio

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
1	5.14%	5.10%	5.23%	6.57%	11.60%	3.87%	21.95%	86.64%	4.62%
2	5.61%	5.34%	4.95%	6.66%	6.23%	5.43%	2.48%	5.04%	53.61%
3	5.22%	5.31%	5.26%	6.18%	5.31%	4.14%	5.35%	3.95%	5.12%
4	5.20%	5.18%	5.37%	6.41%	5.53%	3.88%	5.20%	83.87%	86.32%
5	5.12%	5.06%	5.06%	5.72%	75.21%	3.73%	4.97%	64.35%	73.26%
6	5.51%	5.43%	5.31%	6.09%	5.22%	5.29%	2.36%	4.66%	89.01%
7	5.30%	5.22%	5.05%	6.25%	5.70%	3.73%	5.12%	34.10%	69.72%
8	5.53%	5.31%	5.43%	6.32%	5.62%	3.74%	2.35%	23.19%	94.98%
9	5.27%	5.25%	5.19%	6.65%	5.44%	5.16%	2.52%	85.95%	92.92%
10	5.78%	5.69%	5.69%	6.84%	6.41%	4.54%	3.49%	86.81%	94.12%
Mean	5.37%	5.29%	5.25%	6.37%	13.23%	4.35%	5.58%	47.86%	66.37%
SD	0.23%	0.18%	0.21%	0.33%	21.86%	0.70%	5.90%	37.22%	34.94%

After the models are built, we apply the models to the testing data to make validations. Except the Log Likelihood results in Table 10 and Table 11, and the coefficient estimation results in Appendix 3, other results displayed here are obtained from the testing data validation.

From Table 12 we notice that all techniques except HBW, LCA2 and LCA3 have a quite low total classification error (around 5%), and low standard deviation (5.90% for HBL3 and less than 1% for others). HBW has quite consistent performance over 8 of the 10 experiments and the performance of these 8 experiments is also around 5%. For LCA2 and LCA3, not only the total classification error is high (47.86% and 66.37%), but also the model performance is inconsistent (standard deviation: 37.22% and 34.94%). We examine for more information in Table 13 and Table 14 about the classification error of both the respondents and nonrespondents.

From Table 13 and Table 14, we notice that all techniques have low error rate in one type of classification, and a high error rate in the other. This is because some techniques assign very high predicted probabilities among the customers, such as LCA2 and LCA3, while other methods assign relatively lower predicted probabilities for the customers. Since there are more nonrespondents in the test data, the LCA2 and LCA3 have the highest classification error although they have the lowest respondents classification error. Moreover, these two techniques have inconsistent classification error because they either assign very high or very low predicted probabilities over the customers.

Table 13. Classification Error of Respondent: Original Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
1	87.13%	86.40%	90.07%	85.66%	73.53%	64.71%	33.46%	2.21%	52.21%
2	86.12%	91.46%	79.72%	88.26%	88.97%	88.26%	26.69%	76.16%	3.91%
3	85.77%	88.01%	89.89%	86.52%	87.64%	66.29%	91.01%	71.16%	43.07%
4	82.99%	84.38%	87.15%	85.07%	86.11%	63.89%	86.46%	4.17%	1.04%
5	89.84%	93.36%	92.97%	89.84%	15.63%	62.11%	92.97%	11.72%	25.00%
6	91.32%	91.32%	90.57%	90.19%	99.99%	92.45%	18.11%	77.74%	27.17%
7	84.89%	86.33%	86.33%	83.81%	87.77%	55.40%	86.69%	22.66%	0.72%
8	88.77%	93.33%	89.82%	87.72%	87.72%	58.95%	31.93%	25.61%	40.70%
9	87.86%	88.21%	91.79%	87.14%	90.36%	91.79%	30.36%	14.64%	43.93%
10	84.67%	94.08%	89.90%	86.06%	93.03%	66.90%	37.98%	3.14%	0.00%
Mean	86.94%	89.69%	88.82%	87.03%	81.08%	71.07%	53.57%	30.92%	23.78%
SD	2.56%	3.45%	3.74%	2.03%	23.92%	14.09%	31.21%	31.45%	20.80%

Table 14. Classification Error of Nonrespondent: Original Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
1	0.63%	0.63%	0.57%	2.22%	8.19%	0.53%	21.31%	91.28%	2.00%
2	0.96%	0.37%	0.64%	1.95%	1.46%	0.66%	1.09%	0.94%	56.47%
3	0.85%	0.83%	0.67%	1.83%	0.85%	0.77%	0.71%	0.30%	3.06%
4	0.61%	0.51%	0.55%	1.78%	0.78%	0.35%	0.41%	88.56%	91.35%
5	0.73%	0.49%	0.51%	1.36%	78.30%	0.71%	0.40%	67.08%	75.76%
6	0.88%	0.79%	0.71%	1.55%	0.10%	0.59%	1.51%	0.71%	92.34%
7	0.83%	0.67%	0.48%	1.90%	1.09%	0.83%	0.55%	34.75%	73.59%
8	0.64%	0.14%	0.47%	1.54%	0.80%	0.49%	0.62%	23.05%	98.17%
9	0.61%	0.56%	0.30%	2.10%	0.65%	0.26%	0.95%	89.97%	95.68%
10	1.11%	0.45%	0.70%	2.15%	1.28%	0.85%	1.44%	91.77%	99.69%
Mean	0.78%	0.54%	0.56%	1.84%	9.35%	0.60%	2.90%	48.84%	68.81%
SD	0.17%	0.20%	0.13%	0.29%	24.34%	0.20%	6.48%	40.93%	37.43%

In order to examine whether the high proportion of nonrespondents in the training influences the results, we conduct 10 more experiments using the balanced ratio training data and Tables 15 to 17 show the results.

Table 15. Total Classification Error: Training Data with Balanced Ratio

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
11	30.64%	27.34%	29.47%	35.91%	32.50%	22.78%	22.62%	13.00%	8.44%
12	26.79%	22.74%	26.62%	33.74%	32.42%	17.13%	22.16%	26.68%	27.32%
13	28.22%	22.58%	25.60%	31.63%	33.30%	24.93%	17.21%	79.06%	70.78%
14	28.42%	28.42%	29.69%	34.37%	31.34%	17.54%	19.65%	70.71%	85.72%
15	31.01%	26.74%	29.82%	34.59%	35.73%	26.01%	17.56%	15.07%	33.38%
16	32.57%	25.25%	26.66%	33.58%	10.45%	20.32%	20.79%	12.96%	33.00%
17	31.18%	25.33%	27.28%	34.75%	31.47%	19.94%	12.41%	83.79%	94.63%
18	31.17%	27.86%	31.48%	33.09%	35.56%	23.89%	32.55%	12.32%	94.36%
19	29.23%	26.52%	29.81%	35.19%	32.63%	25.36%	15.08%	80.77%	89.84%
20	28.26%	24.89%	29.20%	35.08%	30.68%	19.77%	17.80%	94.29%	80.05%
Mean	29.75%	25.77%	28.56%	34.19%	30.61%	21.77%	19.78%	48.87%	61.75%
SD	1.82%	2.00%	1.88%	1.23%	7.28%	3.25%	5.47%	35.33%	32.63%

Table 16. Classification Error of Respondents: Balanced Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
11	31.25%	31.25%	30.51%	38.97%	37.13%	23.90%	19.49%	60.29%	40.07%
12	36.30%	37.37%	35.94%	35.94%	35.94%	27.76%	19.22%	54.80%	11.39%
13	36.70%	37.83%	34.08%	37.83%	41.57%	22.10%	25.47%	8.24%	0.37%
14	27.43%	27.43%	31.94%	30.56%	31.60%	22.22%	21.53%	11.46%	0.35%
15	30.47%	30.86%	33.20%	37.11%	36.33%	25.39%	23.05%	48.44%	51.95%
16	35.09%	35.09%	36.23%	36.60%	83.40%	25.28%	21.13%	73.96%	72.08%
17	26.62%	34.53%	33.45%	38.49%	39.21%	32.37%	26.26%	0.00%	0.00%
18	28.42%	27.72%	31.93%	29.47%	35.44%	18.25%	22.81%	35.79%	0.70%
19	30.36%	31.07%	30.71%	38.93%	36.43%	28.21%	18.21%	5.36%	0.00%
20	36.59%	35.89%	37.63%	36.24%	34.84%	26.48%	18.12%	0.00%	6.97%
Mean	31.92%	32.90%	33.56%	36.01%	41.19%	25.20%	21.53%	29.83%	18.39%
SD	3.94%	3.77%	2.41%	3.35%	15.06%	3.91%	2.87%	28.01%	26.45%

From Table 15, we notice that all techniques except LCA2 and LCA3 have similar predictive accuracy (around 30%) and consistent model performance, among which HBL2 and HBL3 has relatively lower error rate (around 20%). LCA2 and LCA3 have high total classification error and the model performance is inconsistent. Again, we look further into the classification error of respondent and nonrespondent.

Table 17. Classification Error of Nonrespondents: Balanced Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
11	27.12%	30.60%	29.41%	35.74%	32.24%	22.71%	22.80%	10.40%	6.69%
12	26.25%	21.90%	26.08%	33.61%	32.22%	16.52%	22.33%	25.06%	28.24%
13	27.76%	21.75%	25.14%	31.29%	32.85%	25.08%	16.76%	82.89%	74.59%
14	28.48%	28.48%	29.56%	34.59%	31.32%	17.27%	19.54%	74.20%	90.75%
15	31.04%	26.52%	29.64%	34.46%	35.70%	26.04%	17.27%	13.34%	32.42%
16	32.44%	24.72%	26.15%	33.41%	6.52%	20.06%	20.77%	9.67%	30.89%
17	31.44%	24.81%	26.93%	34.54%	31.03%	19.24%	11.63%	88.49%	99.94%
18	31.33%	27.87%	31.45%	33.31%	35.57%	24.22%	33.12%	10.94%	99.86%
19	29.17%	26.27%	29.76%	34.98%	32.42%	25.20%	14.91%	85.03%	94.92%
20	27.77%	24.24%	28.70%	35.01%	30.43%	19.37%	17.78%	99.88%	84.38%
Mean	29.28%	25.72%	28.28%	34.10%	30.03%	21.57%	19.69%	49.99%	64.27%
SD	2.13%	2.82%	2.06%	1.25%	8.45%	3.50%	5.81%	38.80%	35.64%

As shown in Table 16, HBL2 (25.20%), HBL3 (21.53%), LCA2 (29.83%), and LCA3 (18.39%) have lower classification error of respondents than other methods (more than 30%). However, the standard deviation of LCA2 (28.01%) and LCA3 (26.45%) is much larger than those of HBL2 (3.91%), HBL3 (2.87%), and other methods (around 3% except HBW).

For classification error of nonrespondents, LCA2 (49.99%) and LCA3 (64.27%) have larger error rate than other methods (less than 35%). Besides, LCA2 (38.80%) and LCA3 (35.64%) have less consistency of classification error than other methods (less than 10%).

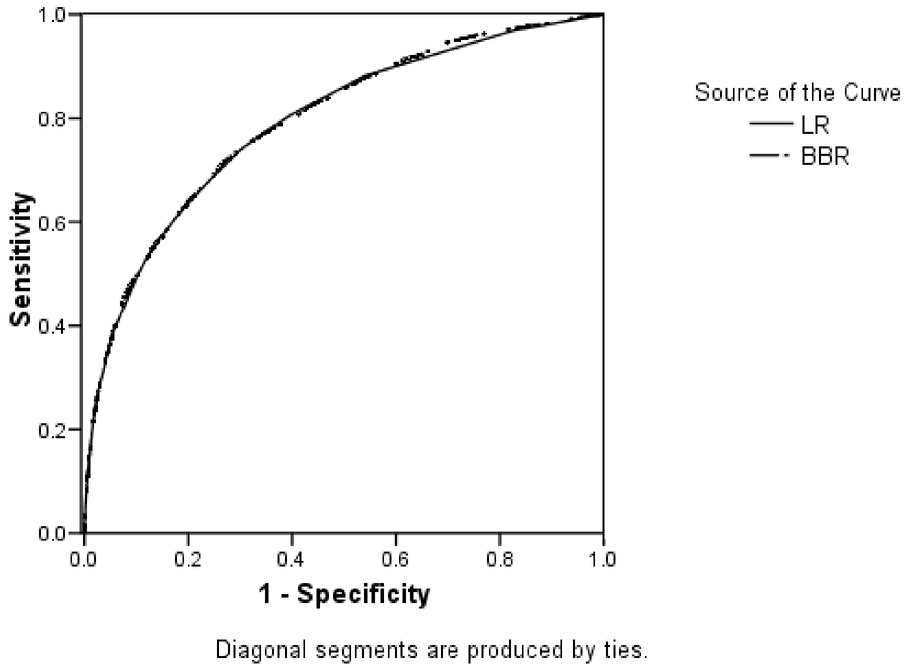
The classification error results of both original and balanced ratio training data show that all methods except LCA2 and LCA3 have relative lower classification error rate and they perform more consistently than LCA2 and LCA3. We check up the predicted probability of the customers and notice that LCA2 and LCA3 either assign very low or very high predicted probabilities over all customers, which means these two methods can only have a better predictive accuracy in either respondent or nonrespondent classification. Since there is much more nonrespondents than respondents in the testing data, LCA2 and LCA3 have very low total classification error when they assign high predicted probabilities to the customers. A possible reason for the inconsistency of model performance of LCA is the latent group memberships. Tests show that LCA does not lead to the same group membership and predicted probability when using the same training data for several times. We can not guarantee the latent group memberships are similar among all experiments.

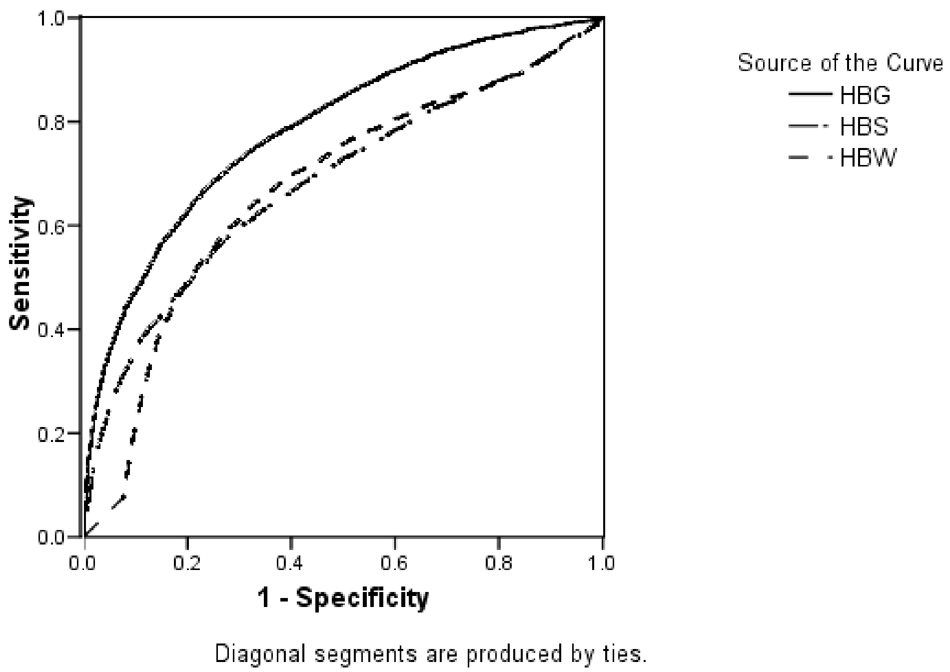
The results also indicate that the proportion of respondent in the training data does not influence the relative model predictive accuracy and model consistency of these techniques. It does have impact on the classification error rate, in which the techniques have less classification error when they using original ratio training data than using balanced ratio training data.

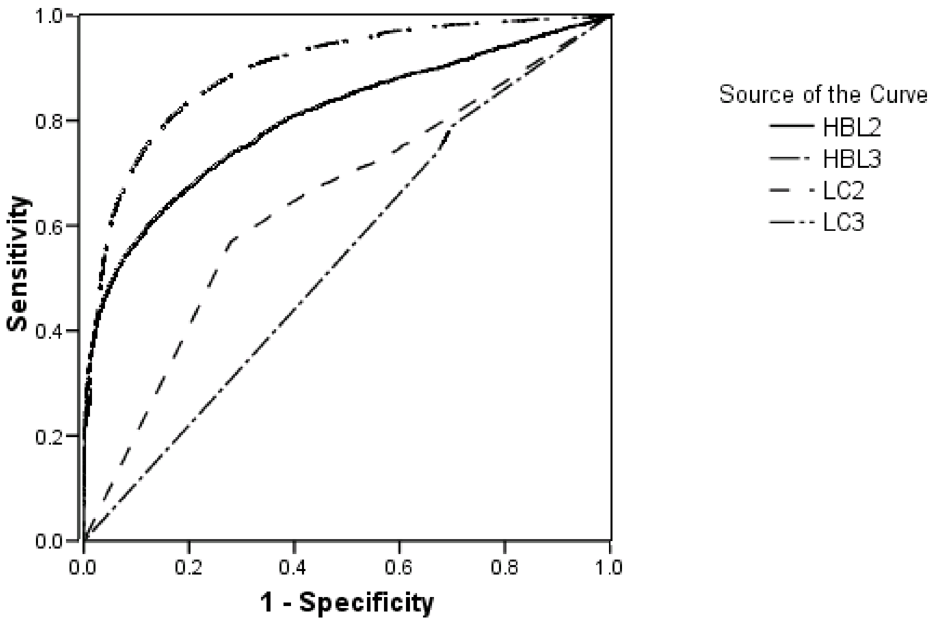
4.4.2 ROC Curve

The simple classification error rate is not enough for model performance assessment, because the cost of assigning a false negative is much greater than assigning a false positive. We compare the model predictive accuracy using the ROC curve. Figure 8 and Figure 9 show the ROC curve of the models and Table 18 and Table 19 indicate the area under the ROC curve.

Figure 8. ROC Curve: Original Ratio Training Data





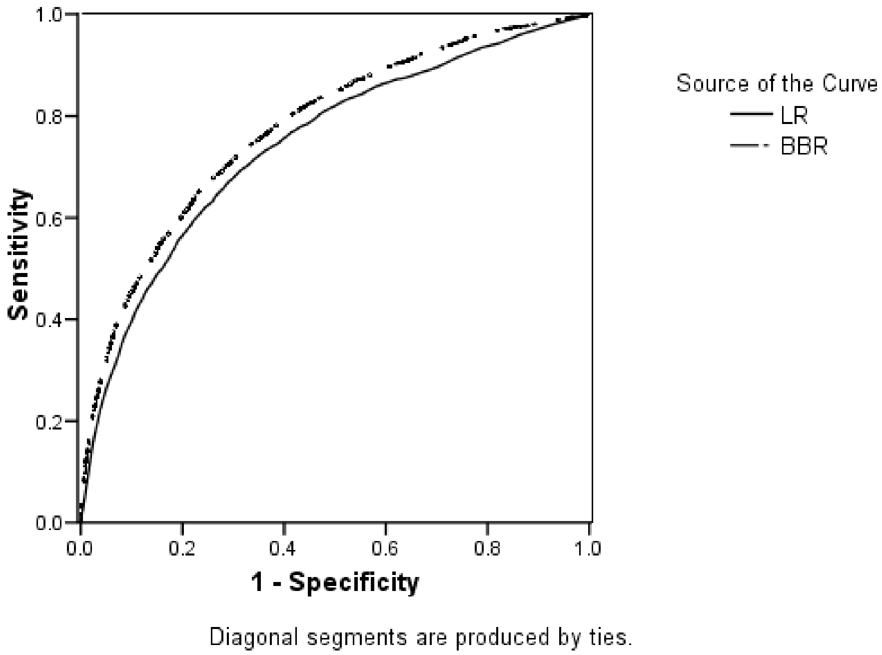


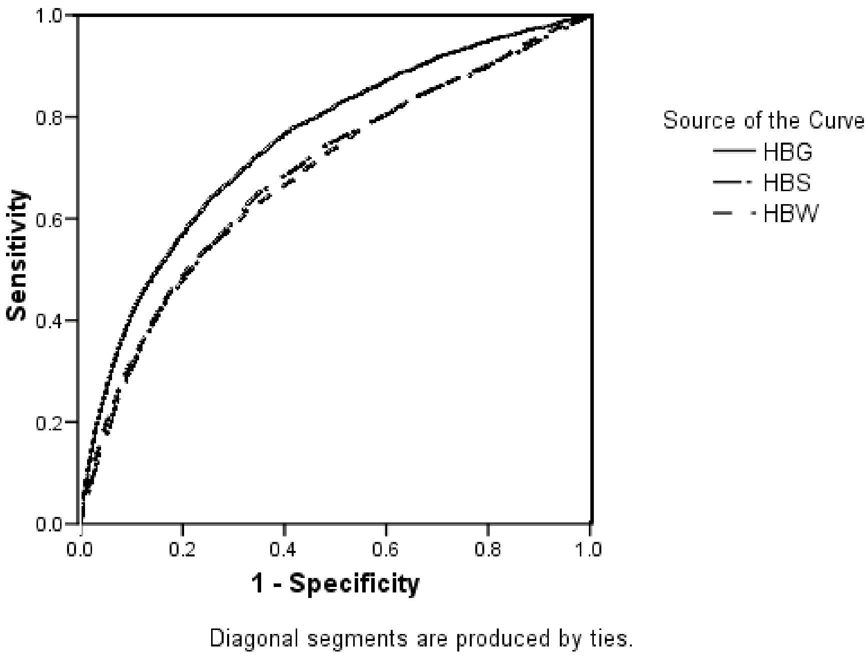
Diagonal segments are produced by ties.

Table 18. Area Under ROC Curve: Original Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
1	0.794	0.796	0.787	0.648	0.703	0.815	0.822	0.610	0.735
2	0.773	0.788	0.783	0.696	0.748	0.794	0.955	0.677	0.710
3	0.797	0.799	0.796	0.722	0.761	0.839	0.789	0.668	0.807
4	0.819	0.822	0.803	0.723	0.781	0.795	0.820	0.561	0.545
5	0.778	0.782	0.762	0.640	0.531	0.787	0.757	0.740	0.495
6	0.767	0.771	0.765	0.668	0.643	0.771	0.975	0.632	0.380
7	0.788	0.790	0.781	0.696	0.748	0.830	0.784	0.812	0.648
8	0.808	0.810	0.814	0.681	0.777	0.850	0.961	0.837	0.304
9	0.810	0.805	0.784	0.662	0.737	0.794	0.965	0.638	0.280
10	0.781	0.791	0.786	0.683	0.732	0.786	0.954	0.562	0.502
Mean	0.792	0.795	0.786	0.682	0.716	0.806	0.878	0.674	0.541
SD	0.017	0.015	0.016	0.028	0.076	0.026	0.090	0.096	0.183

Figure 9. ROC Curve: Balanced Ratio Training Data





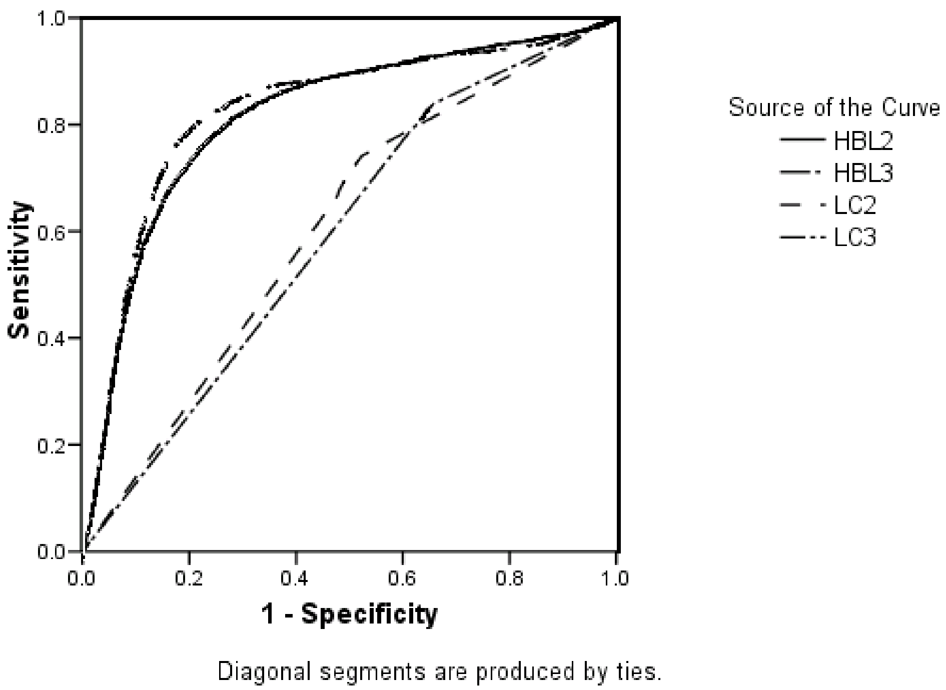


Table 19. Area Under ROC Curve: Balanced Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
11	0.743	0.771	0.762	0.659	0.691	0.819	0.811	0.665	0.794
12	0.738	0.782	0.749	0.682	0.701	0.842	0.822	0.614	0.813
13	0.737	0.760	0.771	0.698	0.674	0.798	0.835	0.543	0.661
14	0.748	0.797	0.762	0.732	0.751	0.828	0.846	0.579	0.557
15	0.742	0.770	0.730	0.642	0.676	0.760	0.812	0.714	0.594
16	0.712	0.775	0.750	0.683	0.687	0.821	0.814	0.593	0.485
17	0.772	0.779	0.761	0.681	0.709	0.810	0.849	0.577	0.501
18	0.757	0.789	0.758	0.716	0.701	0.845	0.779	0.793	0.478
19	0.770	0.792	0.767	0.673	0.705	0.792	0.886	0.555	0.532
20	0.737	0.767	0.712	0.697	0.709	0.834	0.842	0.501	0.543
Mean	0.746	0.778	0.752	0.686	0.700	0.815	0.830	0.613	0.596
SD	0.018	0.012	0.018	0.026	0.022	0.026	0.029	0.088	0.122

The ROC analysis results indicate that HBL2 and HBL3 have larger areas under the ROC curve using both the original ratio (0.806 and 0.878) and balanced ratio training data (0.815 and 0.830) than other techniques (less than 0.8). LCA2 and LCA3 not only have smaller areas (0.674

and 0.541 when using original ratio training data. 0.613 and 0.596 when using balanced ratio training data) under the ROC curve and more inconsistency model performance (Standard deviation 0.096 and 0.183 when using original ratio training data. Standard deviation 0.088 and 0.122 when using balanced ratio training data).

BBR has larger areas under the ROC curve than LR using both original (0.795 vs. 0.792) and balanced (0.778 vs. 0.746) ratio training data. This indicates that both of these methods improve predictive accuracy when more data is available, in which the increase of model performance of LR is more than that of BBR. This certifies that the prior information has more effect on estimation when less data is available.

According to the ROC analysis results, the proportion of respondent in the training data does not have significant influence on relative model performances of these techniques. However, we have to notice the effect of the high proportion of nonrespondents in the testing data. When methods such as LCA2 and LCA3 assign very high predicted probabilities to the customers, these methods improve their true positive rate as well as the false negative rate. Because there are a lot of more nonrespondents in the testing data, the number of false positives greatly exceeds the number of true positives. Thus models may still have smaller area under ROC curve even they have high true positive rate.

4.4.3 Cumulative Lift Value

For direct marketing firms, target customers are those believed to have higher probability of making a purchase when they receive the mail or call. Simply classifying the customers into the respondent or nonrespondent group does not guarantee the success targeting. For example, it is possible that one technique assigns either very high (>=0.5) or very low (<0.5) predicted probabilities to all the customers, and all customers in the first case are considered respondents while nonrespondents in the second case when the cut off value is set to be 0.5. It is unlikely that all “respondents” in the first case are given promotions while no mail is sent to any of the “nonrespondents” in the second case. We then compare the relative importance of those customers by their predicted probability using cumulative lift value. Table 20 to 23 illustrate the cumulative lift value of the first and second deciles of the original and balanced ratio training data respectively.

Table 20. First Decile Cumulative Lift Value: Original Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
1	482.265	474.902	434.406	316.601	279.787	607.433	434.406	589.026	508.035
2	434.501	448.747	434.501	338.341	388.202	445.185	865.440	416.694	683.805
3	461.118	472.365	419.880	374.893	419.880	592.330	453.620	374.893	648.564
4	490.152	500.580	469.294	358.054	420.626	580.534	490.152	597.915	347.625
5	453.561	441.831	422.281	316.711	381.499	570.862	387.091	566.952	449.651
6	419.354	415.576	381.574	313.571	256.902	419.354	918.045	362.685	196.454
7	428.139	442.531	438.933	330.999	410.150	582.845	420.944	590.041	568.454
8	453.514	513.279	499.217	319.921	432.420	611.716	868.356	608.201	267.186
9	464.907	468.483	468.483	339.740	411.264	450.602	876.171	425.569	246.758
10	407.983	446.341	432.393	320.807	369.626	502.133	850.837	509.107	306.859
Mean	449.549	462.464	440.096	332.964	377.036	536.299	656.506	504.108	422.339
SD	26.760	29.427	32.367	20.254	60.700	74.307	233.141	99.282	174.427

From Table 20 and Table 21, we notice that BBR has higher cumulative lift value than LR in the first decile (462.464 vs. 449.549) and similar cumulative lift value in the second decile (303.700 vs. 303.782) which indicates the BBR at least has the same predictive accuracy as LR.

Table 21. Second Decile Cumulative Lift Value: Original Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
1	312.920	309.238	307.398	224.566	230.088	344.212	301.876	323.964	279.787
2	299.165	300.945	284.919	231.496	268.892	288.480	461.212	249.304	400.667
3	309.286	316.784	311.161	258.676	281.169	354.274	309.286	224.936	329.906
4	338.935	328.506	314.601	243.338	300.696	325.030	326.768	340.673	224.218
5	301.071	303.026	297.161	228.736	287.069	328.441	283.476	328.441	265.881
6	266.346	272.013	272.013	239.901	198.343	268.235	483.579	205.899	109.561
7	295.020	298.618	296.819	237.455	269.836	350.786	291.423	345.390	348.988
8	326.952	325.194	332.225	233.788	305.858	351.561	469.334	353.319	144.140
9	302.190	293.249	300.402	225.301	270.004	303.978	473.848	273.580	128.744
10	285.937	289.424	301.629	235.375	266.758	313.833	465.519	298.142	165.634
Mean	303.782	303.700	301.833	235.863	267.871	322.883	386.632	294.365	239.753
SD	20.315	17.095	16.428	10.048	32.212	29.223	89.496	53.108	101.318

Although HBG has higher cumulative lift value than LCA3 in the first decile (440.096 vs. 422.339) and higher cumulative lift value than LCA2 and LCA3 in the second decile (301.833 vs. 294.365, 301.833 vs. 239.753), roughly speaking, the HB using demographic variables has less predictive accuracy than LCA2 and LCA3. This suggests these demographic variables are less appropriate than latent variables when using as clustering variables. Even among the demographic variables, we notice that the cumulative lift values of HBG (440.096 and 301.833), HBS (332.964 and 235.863), and HBW (377.036 and 267.871) differ significantly, which means using different variables for grouping has significant impact on the model performance.

We further notice that, HBL3 (656.506 and 386.632) has higher cumulative lift value than HBL2 (536.299 and 322.833) and both these two techniques have higher cumulative lift values than other methods. This suggests HB using latent variables for classification outperform other techniques, and in this case, the HB using 3 latent classes for classification is the best set. However, we notice that HBL2 (74.307 and 29.223), HBL3 (233.141 and 89.496), LCA2 (99.282 and 53.108) and LCA3 (174.427 and 101.318) have larger standard deviation of cumulative lift value in both first and second deciles. This may be caused by the instability of latent class assignment, because even the same data may lead to different group membership using the same algorithm for several times.

To further check up whether the proportion of respondents in the training data has influence on the model performance, we conduct experiments using balanced ratio training data. Table 22 and Table 23 illustrate the cumulative lift value results.

Table 22. First Decile Cumulative Lift Value: Balanced Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
11	397.592	427.044	404.955	250.336	305.557	449.132	445.451	360.778	611.114
12	359.710	423.816	377.517	267.111	288.480	519.977	316.972	185.197	267.111
13	344.901	419.880	404.884	292.416	262.425	296.165	521.101	618.573	284.918
14	417.150	455.389	424.103	371.959	396.293	507.533	504.057	723.060	472.770
15	316.711	387.091	355.811	230.691	285.431	156.401	574.772	402.731	160.311
16	298.459	385.352	355.129	256.902	230.456	415.576	340.017	260.680	200.232
17	395.759	420.944	388.563	255.445	298.618	503.693	633.215	100.739	514.487
18	362.108	453.514	362.108	312.889	333.983	453.514	351.561	537.889	214.452
19	411.264	461.331	400.535	303.978	293.249	454.178	665.175	611.532	514.974
20	345.217	404.496	292.911	296.398	296.398	585.822	439.367	512.594	707.869
Mean	364.887	423.886	376.652	283.813	299.089	434.199	479.169	431.377	394.824
SD	40.082	26.919	37.542	40.831	43.593	123.882	122.258	204.229	192.468

Table 22 and Table 23 indicate that all techniques have smaller cumulative lift value when using balanced ratio training data than using original ratio training data, but the relative model performances do not vary significantly.

In detail, the BBR (423.886 and 288.248) still has more predictive accuracy than LR (364.887 and 270.129), but the difference in cumulative lift value increases. This result is consistent with Rossi and Allenby's research (2003) that the prior information has more influence when data information is limited.

Table 23. Second Decile Cumulative Lift Value: Balanced Ratio Training Data

Experiment	LR	BBR	HBG	HBS	HBW	HBL2	HBL3	LCA2	LCA3
11	281.630	281.628	268.743	231.929	233.770	340.530	371.822	237.451	333.168
12	267.111	283.138	270.673	224.373	251.085	365.052	341.902	117.529	309.849
13	260.550	279.295	288.667	232.433	219.312	303.663	374.893	324.282	380.516
14	285.053	307.648	286.791	271.148	293.743	389.340	382.388	375.435	319.815
15	246.331	285.431	263.926	207.231	222.871	314.756	383.181	306.936	91.885
16	247.457	275.791	255.013	224.789	204.010	345.684	351.351	202.121	119.006
17	286.026	287.825	277.031	219.466	241.053	329.200	401.156	66.560	422.743
18	281.249	304.100	274.218	240.819	233.788	356.835	298.827	362.108	172.265
19	286.097	300.402	282.520	223.513	241.394	314.706	418.416	327.223	366.561
20	259.784	277.219	254.554	228.401	249.323	357.421	406.240	371.369	373.113
Mean	270.129	288.248	272.214	230.410	239.035	341.719	373.018	269.101	288.892
SD	15.898	11.597	12.053	16.836	24.004	26.596	35.110	109.421	117.411

The cumulative lift values of HBG (376.652 and 272.214), HBS (283.813 and 230.410), and HBW (299.089 and 239.035) in balanced ratio training are much less than those in original ratio training (HBG: 440.096 and 301.833, HBS: 332.964 and 235.863, and HBW: 377.036 and 267.871). One explanation is that, the balanced ratio training data contains insufficient information for classifications with so many groups.

As in the original ratio training data experiments, the HBL2 (434.199 and 341.719) and HBL3 (479.169 and 373.018) have higher cumulative lift values than LCA2 (431.377 and 269.101) and LCA3 (394.824 and 288.892) in both the first and second deciles, and HBL3 (479.169 and 373.018) has more predictive accuracy than HBL2 (434.199 and 341.719). Again, these four techniques have relatively larger standard deviation (HBL2: 123.882 and 26.596, HBL3: 122.258 and 35.110, LCA2: 204.229 and 109.421, and LCA3: 192.468 and 117.411) than other methods (around 40 in the first decile and 20 in the second).

4.4.4 Summary

In sum, according to the results of the experiments, both original ratio and balanced ratio training, we can conclude that, when the customers are analyzed as a whole group, the Bayesian method BBR has more predictive accuracy than classical method such as logistic regression. When customers are separated into different groups, the Bayesian methods using demographics variables have less predictive accuracy than classical methods such as Latent Class Analysis. However, when using the same latent group membership, the Bayesian methods have better model accuracy than Latent Class Analysis.

The BBR has larger cumulative lift value than LR and the difference is more significant in balanced ratio training than original ratio training. This indicates the Bayesian method has more effects when the data is less because it can benefit more from the prior knowledge. When more data is available, the prior knowledge has relatively less influence since more information, such as likelihood, can be derived from the data.

The larger cumulative lift values of HBL2 and HBL3 over the LCA2 and LCA3 in both original ratio and balanced ratio training indicates the Hierarchical Bayes using latent class membership has greater predictive accuracy than Latent Class Analysis in direct marketing customer response prediction. Although these four techniques have large standard deviation of cumulative lift values in the first decile, the standard deviation in the second decile is much less. Such consistent results further support our conclusion about the advantage of Bayesian methods over other methods in the direct marketing customer response prediction.

Although some researchers argue that the aggregate analysis has the same predictive accuracy as Hierarchical Bayes and Latent Class Analysis (Natter and Feurstein, 2002), our results show that, in direct marketing customer prediction where the ultimate decision is made in an individual level, the Hierarchical Bayes is more appropriate than aggregate Bayesian method.

5 Conclusion

5.1 Findings

Our research is one of the first empirical studies of Bayesian applications in direct marketing customer response prediction. The results support the importance of prior information, the unique feature of Bayesian methods, in direct marketing modeling where there is a lot of transaction history. Among all variables available in the direct marketing data, RFM variables are the most important in customer response predictions, thus when setting prior variances, these variables should be assigned relatively smaller values. This means that, the RFM variables do not vary significantly in influencing the posterior estimations.

Accounting for customer heterogeneity is important in marketing modeling, especially in direct marketing context where firms need to adopt strategies toward individual customers. Our results show the advantage of using Hierarchical Bayes over aggregate Bayesian methods, which means the customers in direct marketing context are heterogeneous and they should be treated differently. However, when comparing with Latent Class Analysis, the Hierarchical Bayes not only results in more predictive accuracy. But the HB model has more consistent model performance. In Hierarchical Bayes models, the coefficient estimation of different groups of customers is similar in some extent while Latent Class Analysis builds different models for different latent groups. This indicates that the continuous heterogeneity is more applicable in direct marketing customer response prediction than discrete heterogeneity.

Bayesian methods are free from assumptions such as homogeneity and discrete heterogeneity among customers. Besides, the empirical direct marketing data used in this thesis contain a large number of observations as well as independent variables. Although sample selection and variable selection processes are conducted, the training data are still very large. With the development of computational methods, many computational problems have been solved. However, our experiments indicate that the Bayesian methods not only spend less time in model constructing, but also are capable in finishing certain type of task that can not be accomplished by other methods. For example, Hierarchical Bayes is able to conduct an analysis with 51 groups which can not be accomplished by the Latent Class Analysis.

Our study also shows that, in direct marketing customer response prediction, the cumulative lift value is better than simple classification error and ROC analysis as a validation criterion. The simple classification error does not consider the relative importance of false negative and false positive. The ROC analysis is not effective enough when there is much more nonrespondents in the testing data. The cumulative lift value considers customers by their predicted probabilities. Such a relative importance of customers is very important for direct marketing firms to make the ranking and identify target customers.

Moreover, our study using both original and balanced ratio training data suggests that the proportion of respondents in the training data does not have significant impact on the relative model performances of different techniques. The original ratio training data where more information are provided are able to help those techniques to have better understanding of the customers and make more accurate estimations.

The test results further support the applicability of Bayesian methods in direct marketing context, especially in customer response prediction. However, one cannot assume the universal applicability of the Bayesian methods since no single method can dominate other techniques in the model performance. When analyzing data with adequate historical information and where customer heterogeneity is assumed, the Bayesian methods represent a strong alternative method for modeling customer responses to direct marketing promotions.

5.2 Managerial Implications

Our study testifies the applicability of Bayesian methods in direct marketing customer response prediction. The results of the empirical study indicate that the Bayesian methods, by incorporating the prior information of customers, have greater predictive power than other classical direct marketing techniques. Using Bayesian methods can potentially improve the model accuracy and help increase the customer response rate. Most importantly, the better performance of Bayesian methods should convince direct marketing firms of the importance of historical data in estimating future customer behaviors. Firms should pay more attention to variables such as RFM and understand their effects.

Since the prior information is more important when less data are provided, direct marketing firms do not need to include many customers into the analysis, which may cause difficulties in computation. They can make good use of the prior information, both subjective priors and objective priors, especially when there is little information available or they have to pay expenses on acquiring additional data.

Our results show that the HB models do not match the performance of latent class analysis. This may be due to the fact that some of the hierarchy variables in this study are zip-code level variables which represent the average information of all customers in these areas. The average of data causes a loss of information. When data privacy is not a concern, companies may use customer-level data to choose more appropriate hierarchy variables, which may provide better opportunities to explore the advantages of the Hierarchical Bayes approach and arrive at more accurate predictions of customer behavior.

Our test results also show that, when companies find it difficult to cluster customers by their demographic or geographic variables, they can consider using latent variable for group membership. However, we have to notice that although such methods may have very high averaged predicted accuracy, the model performance is inconsistent and has very high variance. In such cases, direct marketing firms may choose the methods which may not have the best model accuracy, but they are more stable in model performances.

5.3 Limitations

This study mainly concerns with the response rate of the direct marketing customers. Other factors, such as profit, should be taken into consideration. According to the result of customer response prediction, customers with the highest probability of response are considered more important and specific strategies are provided for them. However, from an economic perspective, these customers may not bring firms as much profit as some other customers with a relatively lower probability of response. Firms should balance between the profit brought through a transaction and the likelihood of the occurrence of the transaction.

Besides, the variable selection process still remains as an unsolved problem which needs to be addressed. Our study selects 109 independent variables from a total of 307. The number of independent variables still seems too large. How to select an appropriate set of independent variables, especially from a large and noisy dataset, is not well understood. Moreover, although a normal setting of prior distribution is considered appropriate in many cases, we did not check the posterior distribution to see whether the normal setting works. The geometric setting of prior variances reduces the efforts in obtaining prior knowledge, but it does not include all possible prior variance values.

In this study, we apply the latent group membership produced by LCA2 and LCA3 into the Hierarchical Bayes estimation. Although more and more studies are using the boosting and bagging method in model building, whether such a method is applicable in direct marketing customer response prediction is not well understood. Hierarchical Bayes may build different latent groups itself. However, we note that the main concern in this study is not the identifying of latent classes. Instead, we examine the model performance of Hierarchical Bayes and Latent Class Analysis under the same condition to discover their differences. The latent variables are proved important in estimation, but since these variables are seldom related with the demographic information, it is very difficult for firms to make applications in new market.

Moreover, the validation method apply in this study is unconventional. The reason we use 10 pairs of train and test data is to reduce the probability of model over fit. Normally speaking, a 10-fold cross validation should be considered in this situation. However, we do not use cross validation because of practical reasons. Although the 10 fold train and validation method gives us some insight of model stability, the theorem behind needs further exploration.

5.4 Directions for Future Research

In this study, the BBR has more predictive accuracy than HB using demographic variables. This is not to suggests that disaggregate estimation does not necessarily lead to better results than those models that do not consider customer heterogeneity. Instead of using demographic variables at the zip code level, future studies should consider use individual level information. Hierarchical Bayes models using such data may render better performance in predicting customer responses. On the other hand, the nature of consumer heterogeneity warrants further investigation. Future studies may compare the performance of models that use hierarchical variables and with those that treat such heterogeneity as unobserved using the latent class approach.

Another important issue to be addressed is the trade-off between the model accuracy and variance. As we discussed in the Data Analysis part, the HBL2 and HBL3 have the greatest predictive accuracy, but the models built by these two methods are quite inconsistent and exhibit greater variance. Other methods such as BBR also has very good performance and at the same more consistent model performance. The issue of how should researchers and practitioners choose from these alternative models needs further discussion.

Lastly, the Hierarchical Bayes model represents a revolutionary method that can potentially improve the accuracy at which we understand and predict consumer behavior. As the No Free Lunch axiom implies, no single method has all the solutions to the various problems in empirical research. An increasing repertoire of techniques has been applied to data mining with direct marketing data. As alternative methods may offer better solutions to specific problems encountered in empirical analyses, researchers may consider combine the strengths of these methods using data mining procedures such as boosting and bagging, and can potentially build more powerful predictive models when such methods can complement each other.

The Latent Class Analysis does not have a table performance in this study, this may because of the number of latent groups chosen, number of iterations, and some other parameter settings.

Different settings should be considered in future study to examine whether the Latent Class Analysis will have a better performance in direct marketing customer response prediction.

Appendix 1. Variables Description

Variables	Data Description
targact	Ordered from Prom 85 in Targ Wndw (Y/N)
ord101	Order Yr 1, Prom 01 (Y/N)
ord117	Order Yr 1, Prom 17 (Y/N)
ord121	Order Yr 1, Prom 21 (Y/N)
ord150	Order Yr 1, Prom 50 (Y/N)
ord165	Order Yr 1, Prom 65 (Y/N)
ord172	Order Yr 1, Prom 72 (Y/N)
ord185	Order Yr 1, Prom 85 (Y/N)
ord193	Order Yr 1, Prom 93 (Y/N)
ord201	Order Yr 2, Prom 01 (Y/N)
ord217	Order Yr 2, Prom 17 (Y/N)
ord221	Order Yr 2, Prom 21 (Y/N)
ord250	Order Yr 2, Prom 50 (Y/N)
ord265	Order Yr 2, Prom 65 (Y/N)
ord272	Order Yr 2, Prom 72 (Y/N)
ord285	Order Yr 2, Prom 85 (Y/N)
ord293	Order Yr 2, Prom 93 (Y/N)
ord301	Order Yr 3, Prom 01 (Y/N)
ord317	Order Yr 3, Prom 17 (Y/N)
ord321	Order Yr 3, Prom 21 (Y/N)
ord350	Order Yr 3, Prom 50 (Y/N)
ord365	Order Yr 3, Prom 65 (Y/N)
ord372	Order Yr 3, Prom 72 (Y/N)
ord385	Order Yr 3, Prom 85 (Y/N)
ord393	Order Yr 3, Prom 93 (Y/N)
ord401	Order Yr 4, Prom 01 (Y/N)
ord411	Order Yr 4, Prom 11 (Y/N)
ord417	Order Yr 4, Prom 17 (Y/N)
ord421	Order Yr 4, Prom 21 (Y/N)
ord424	Order Yr 4, Prom 24 (Y/N)
ord435	Order Yr 4, Prom 35 (Y/N)
ord441	Order Yr 4, Prom 41 (Y/N)
ord445	Order Yr 4, Prom 45 (Y/N)
ord450	Order Yr 4, Prom 50 (Y/N)
ord457	Order Yr 4, Prom 57 (Y/N)
ord464	Order Yr 4, Prom 64 (Y/N)
ord465	Order Yr 4, Prom 65 (Y/N)
ord470	Order Yr 4, Prom 70 (Y/N)

ord472	Order Yr 4, Prom 72 (Y/N)
ord480	Order Yr 4, Prom 80 (Y/N)
ord482	Order Yr 4, Prom 82 (Y/N)
ord485	Order Yr 4, Prom 85 (Y/N)
ord495	Order Yr 4, Prom 95 (Y/N)
ord493	Order Yr 4, Prom 93 (Y/N)
ord400	Order Yr 4, Prom 00 (Y/N)
fstycls5	Year of First Product Class 5 Order
fstycls2	Year of First Product Class 2 Order
fstycls7	Year of First Product Class 7 Order
fstycls6	Year of First Product Class 6 Order
fstycls1	Year of First Product Class 1 Order
fstycls3	Year of First Product Class 3 Order
fstycls4	Year of First Product Class 4 Order
fstcls5	First Order Product Class 5
fstcls2	First Order Product Class 2
fstcls7	First Order Product Class 7
fstcls6	First Order Product Class 6
fstcls1	First Order Product Class 1
fstcls3	First Order Product Class 3
fstcls4	First Order Product Class 4
lstycls5	Year of Last Product Class 5 Order
lstycls2	Year of Last Product Class 2 Order
lstycls7	Year of Last Product Class 7 Order
lstycls6	Year of Last Product Class 6 Order
lstycls1	Year of Last Product Class 1 Order
lstycls3	Year of Last Product Class 3 Order
lstycls4	Year of Last Product Class 4 Order
lstcls5	Last Order Product Class 5 (Y/N)
lstcls2	Last Order Product Class 2 (Y/N)
lstcls7	Last Order Product Class 7 (Y/N)
lstcls6	Last Order Product Class 6 (Y/N)
lstcls1	Last Order Product Class 1 (Y/N)
lstcls3	Last Order Product Class 3 (Y/N)
lstcls4	Last Order Product Class 4 (Y/N)
yrord1	Orders in Yr 1
yrord2	Orders in Yr 2
yrord3	Orders in Yr 3
yrord4	Orders in Yr 4
yrord5	Orders in Yr 5
yrord6	Orders in Yr 6
yrord7	Orders in Yr 7
yrord8	Orders in Yr 8

yrord9	Orders in Yr 9
yrord10	Orders in Yr 10
yrord11	Orders in Yr 11
yrord12	Orders in Yr 12
prord01	Lifetime Orders Prom 01
prord03	Lifetime Orders Prom 03
prord05	Lifetime Orders Prom 05
prord07	Lifetime Orders Prom 07
prord08	Lifetime Orders Prom 08
prord10	Lifetime Orders Prom 10
prord11	Lifetime Orders Prom 11
prord12	Lifetime Orders Prom 12
prord17	Lifetime Orders Prom 17
prord21	Lifetime Orders Prom 21
prord24	Lifetime Orders Prom 24
prord35	Lifetime Orders Prom 35
prord41	Lifetime Orders Prom 41
prord45	Lifetime Orders Prom 45
prord50	Lifetime Orders Prom 50
prord57	Lifetime Orders Prom 57
prord58	Lifetime Orders Prom 58
prord60	Lifetime Orders Prom 60
prord63	Lifetime Orders Prom 63
prord64	Lifetime Orders Prom 64
prord65	Lifetime Orders Prom 65
prord70	Lifetime Orders Prom 70
prord72	Lifetime Orders Prom 72
prord80	Lifetime Orders Prom 80
prord82	Lifetime Orders Prom 82
prord85	Lifetime Orders Prom 85
prord87	Lifetime Orders Prom 87
prord95	Lifetime Orders Prom 95
prord93	Lifetime Orders Prom 93
prord00	Lifetime Orders Prom 00
cat1	Purchase CAT 1 (Y/N)
cat2	Purchase CAT 2 (Y/N)
cat3	Purchase CAT 3 (Y/N)
cat4	Purchase CAT 4 (Y/N)
cat5	Purchase CAT 5 (Y/N)
cat6	Purchase CAT 6 (Y/N)
cat7	Purchase CAT 7 (Y/N)
cat11	Purchase CAT 11 (Y/N)
cat13	Purchase CAT 13 (Y/N)

cat15	Purchase CAT 15 (Y/N)
cat18	Purchase CAT 18 (Y/N)
cat22	Purchase CAT 22 (Y/N)
cat23	Purchase CAT 23 (Y/N)
cat24	Purchase CAT 24 (Y/N)
cat25	Purchase CAT 25 (Y/N)
cat26	Purchase CAT 26 (Y/N)
cat31	Purchase CAT 31 (Y/N)
cat32	Purchase CAT 32 (Y/N)
cat33	Purchase CAT 33 (Y/N)
cnvcat1	First Purchase CAT 1 (Y/N)
cnvcat2	First Purchase CAT 2 (Y/N)
cnvcat3	First Purchase CAT 3 (Y/N)
cnvcat4	First Purchase CAT 4 (Y/N)
cnvcat5	First Purchase CAT 5 (Y/N)
cnvcat6	First Purchase CAT 6 (Y/N)
cnvcat7	First Purchase CAT 7 (Y/N)
cnvcat11	First Purchase CAT 11 (Y/N)
cnvcat13	First Purchase CAT 13 (Y/N)
cnvcat15	First Purchase CAT 15 (Y/N)
cnvcat18	First Purchase CAT 18 (Y/N)
cnvcat22	First Purchase CAT 22 (Y/N)
cnvcat23	First Purchase CAT 23 (Y/N)
cnvcat24	First Purchase CAT 24 (Y/N)
cnvcat25	First Purchase CAT 25 (Y/N)
cnvcat26	First Purchase CAT 26 (Y/N)
cnvcat31	First Purchase CAT 31 (Y/N)
cnvcat32	First Purchase CAT 32 (Y/N)
cnvcat33	First Purchase CAT 33 (Y/N)
fphone	First order by phone (Y/N)
lphone	Last order phone (Y/N)
convprom	Conversion Prom
salcat	Dollar class of lifetime avg ord
salflg	Dollar class of avg ord last year
rlpolk	On Polk file (Y/N)
tande	Used T&E Cred Card (Y/N)
crdcd	Used Cred Card (Y/N)
hcrd	Used House Credit Card (Y/N)
cash	Cash Order (Y/N)
tele	Bought Telemktng Prom (Y/N)
bus	Business Customer (Y/N)
recmon	Months since last order
totsal1	Sales in Yr 1

totsal2	Sales in Yr 2
totsal3	Sales in Yr 3
totsal4	Sales in Yr 4
totsal5	Sales in Yr 5
totsal6	Sales in Yr 6
totsal7	Sales in Yr 7
totsal8	Sales in Yr 8
totsal9	Sales in Yr 9
totsal10	Sales in Yr 10
totsal11	Sales in Yr 11
totsal12	Sales in Yr 12
ordcls5	5 Year Product Class 5 Orders
ordcls2	5 Year Product Class 2 Orders
ordcls7	5 Year Product Class 7 Orders
ordcls6	5 Year Product Class 6 Orders
ordcls1	5 Year Product Class 1 Orders
ordcls3	5 Year Product Class 3 Orders
ordcls4	5 Year Product Class 4 Orders
salcls5	5 Year Sales Product Class 5
salcls2	5 Year Sales Product Class 2
salcls7	5 Year Sales Product Class 7
salcls6	5 Year Sales Product Class 6
salcls1	5 Year Sales Product Class 1
salcls3	5 Year Sales Product Class 3
salcls4	5 Year Sales Product Class 4
sordcls5	12 Year Orders Product Class 5
sordcls2	12 Year Orders Product Class 2
sordcls7	12 Year Orders Product Class 7
sordcls6	12 Year Orders Product Class 6
sordcls1	12 Year Orders Product Class 1
sordcls3	12 Year Orders Product Class 3
sordcls4	12 Year Orders Product Class 4
ssorcls5	12 Year Sales Product Class 5
ssorcls2	12 Year Sales Product Class 2
ssorcls7	12 Year Sales Product Class 7
ssorcls6	12 Year Sales Product Class 6
ssorcls1	12 Year Sales Product Class 1
ssorcls3	12 Year Sales Product Class 3
ssorcls4	12 Year Sales Product Class 4
crcpr01	Circs Prom 01
crcpr10	Circs Prom 10
crcpr17	Circs Prom 17
crcpr21	Circs Prom 21

crcpr24	Circs Prom 24
crcpr35	Circs Prom 35
crcpr50	Circs Prom 50
crcpr60	Circs Prom 60
crcpr65	Circs Prom 65
crcpr72	Circs Prom 72
crcpr85	Circs Prom 85
crcpr93	Circs Prom 93
convsale	Conversion Sales
aocls5	12 Yr Avg Ord, Product Class 5
aocls2	12 Yr Avg Ord, Product Class 2
aocls7	12 Yr Avg Ord, Product Class 7
aocls6	12 Yr Avg Ord, Product Class 6
aocls1	12 Yr Avg Ord, Product Class 1
aocls3	12 Yr Avg Ord, Product Class 3
aocls4	12 Yr Avg Ord, Product Class 4
totord	Lifetime Orders
totsale	Lifetime Sales
custaord	Lifetime Average Order
maxdol	Largest Order
rec1	Recency of last Order< 7 months (Y/N)
rec2	Recency of last Order 7-12 months (Y/N)
rec3	Recency of last Order 13-18 months (Y/N)
rec4	Recency of last Order 19-24 months (Y/N)
rec5	Recency of last Order 25-30 months (Y/N)
rec6	Recency of last Order 31-36 months (Y/N)
rec7	Recency of last Order > 36 months (Y/N)
ageyngtl	Age of Youngest Tradeline
aveagetl	Ave Age of All Active + Paid Tradelines
avageotl	Ave Age of All Open+Active Tradelines
totnumtl	Total Number of All Tradelines
aveotlbl	Ave Balance of All Open+Active TLs
numsattl	Num of All TLs with Satisfactory Ratings
numdertl	Num of All TLs with Derogatory Ratings
opntl1yr	Num of All TLs Opened in the Last 12 Mo
autoloan	Highest Ave Loan Amt for Auto Acct Types
numbcact	Number of All Bank Card Accounts
avebcbal	Ave Balance of All Open+Act Bank Cards
avebccl	Ave Credit Lim for All Active Bank Cards
highbccl	Highest Cred Lim for All Act Bank Cards
numsatbc	Num of All Bank Cards w. Satisfac Rating
ageoldbc	Age of Oldest Bank Card Line
opnbc1yr	Num Bank Cards Opened in Last 12 Months

numccact	Number of All Credit Card Accounts
aveccbal	Ave Balance of All Open+Act Credit Cards
avecccl	Ave Cred Lim for All Open+Act Cred Cards
ageoldcc	Age of Oldest Credit Card Account
opncc1yr	Num of Credit Cards Opened in Last 12 Mo
nowcurtl	Num of All TLs Past Delinq + Now Current
tl30dayr	Num of All TLs Once 30 Days Late
mult30dr	Num of All TLs 2x or More 30 Days Late
tl60dayr	Num of All TLs Once 60 Days Late
tl90dayr	Num of All TLs Once 90 Days Late
tlgt90dr	Num of All TLs Once > 90 Days Late
tldelinq	Num of All TLs with Serious Delinquency
numflitl	Num of TLs from Financial Lending Inst.
numinqry	Number of Inquiries by Credit Grantors
numtl4mo	Num of Active TLs w. Bal>0 and Age<4 Mo
numtl13m	Num of Active TLs w. Bal>0 and Age<13 Mo
retailtl	Num of All Retail Credit Tradelines
avertbal	Ave Balance of All Active Retail TLs
averetcl	Ave Cred Lim for All Active Retail TLs
hghretcl	Highest Cred Lim for All Act Retail TLs
ageoldrt	Age of Oldest Retail Credit Tradeline
opnret1y	Num of Retail TLs Opened in Last 12 Mo
dmawlthr	DMA wealth rating
incmindx	income index
wealthrt	wealth rating
prcwhte	% occupied HH-white
prcblck	% occupied HH-black
prchisp	% occupied HH-Hispanic
prcun18	% HH with 1+ under 18
prcowno	% OOHH
prcthre	% HH-3+ persons
perperhh	persons per HH
prcncd1	% NCDB HH-1 unit structures
medschyr	median years school for people age 25+
prc25ba	% population age 25+ with a BA or more
prcncd3	% NCDB HH-3+ unit structures
prcnc10	% NCDB HH-10+ unit structures
oomedhvl	OOH census median home value in 000s
oohvi	OOH home value index
prcoohv	OOH home value percentile
ispsa	index social position small areas
prcrent	% occupied housing unit-renter occupied
prc3544	% occupied housing unit-age 35-44

prc4554	% occupied housing unit-age 45-54
prc5564	% occupied housing unit-age 55-64
prc65p	% occupied housing unit-age 65+
prc55p	% occupied housing unit age 55+
hhmedage	householders' median age
cemi	current estimated median income in 000s
prc500k	% OOH value $500,000+
prc200k	% OOH value $200,000+
prc100k	% OOH value $100,000+
prchhfm	% HH that are families
populat	population

Appendix 2. List of Variables Chosen for Data Analysis

Variables	Data Description
targact	Ordered from Prom 85 in Target Window (Y/N)
ord165	Order Yr 1, Prom 65 (Y/N)
ord172	Order Yr 1, Prom 72 (Y/N)
ord185	Order Yr 1, Prom 85 (Y/N)
ord193	Order Yr 1, Prom 93 (Y/N)
ord201	Order Yr 2, Prom 01 (Y/N)
ord250	Order Yr 2, Prom 50 (Y/N)
ord285	Order Yr 2, Prom 85 (Y/N)
ord301	Order Yr 3, Prom 01 (Y/N)
ord317	Order Yr 3, Prom 17 (Y/N)
ord350	Order Yr 3, Prom 50 (Y/N)
ord385	Order Yr 3, Prom 85 (Y/N)
ord424	Order Yr 4, Prom 24 (Y/N)
ord435	Order Yr 4, Prom 35 (Y/N)
ord450	Order Yr 4, Prom 50 (Y/N)
ord485	Order Yr 4, Prom 85 (Y/N)
ord493	Order Yr 4, Prom 93 (Y/N)
fstycls3	Year of First Product Class 3 Order
fstcls2	First Order Product Class 2
lstycls3	Year of Last Product Class 3 Order
yrord1	Orders in Yr 1
yrord2	Orders in Yr 2
yrord3	Orders in Yr 3
yrord4	Orders in Yr 4
yrord5	Orders in Yr 5
yrord6	Orders in Yr 6
yrord7	Orders in Yr 7
yrord8	Orders in Yr 8
yrord9	Orders in Yr 9
yrord10	Orders in Yr 10
yrord11	Orders in Yr 11
yrord12	Orders in Yr 12
prord12	Lifetime Orders Prom 12
prord80	Lifetime Orders Prom 80
prord85	Lifetime Orders Prom 85
cat18	Purchase CAT 18 (Y/N)
cat25	Purchase CAT 25 (Y/N)
cat26	Purchase CAT 26 (Y/N)

cat31	Purchase CAT 31 (Y/N)
cat33	Purchase CAT 33 (Y/N)
cnvcat22	First Purchase CAT 22 (Y/N)
cnvcat25	First Purchase CAT 25 (Y/N)
cnvcat26	First Purchase CAT 26 (Y/N)
cnvcat33	First Purchase CAT 33 (Y/N)
fphone	First order by phone (Y/N)
convprom	Conversion Prom
salcat	Dollar class of lifetime avg ord
salflg	Dollar class of avg ord last year
hcrd	Used House Credit Card (Y/N)
cash	Cash Order (Y/N)
tele	Bought Telemktng Prom (Y/N)
bus	Business Customer (Y/N)
recmon	Months since last order
totsal1	Sales in Yr 1
totsal2	Sales in Yr 2
totsal3	Sales in Yr 3
totsal4	Sales in Yr 4
totsal5	Sales in Yr 5
totsal6	Sales in Yr 6
totsal7	Sales in Yr 7
totsal8	Sales in Yr 8
totsal9	Sales in Yr 9
totsal10	Sales in Yr 10
totsal11	Sales in Yr 11
totsal12	Sales in Yr 12
ordcls2	5 Year Product Class 2 Orders
ordcls7	5 Year Product Class 7 Orders
ordcls3	5 Year Product Class 3 Orders
sordcls6	12 Year Orders Product Class 6
ssorcls5	12 Year Sales Product Class 5
ssorcls6	12 Year Sales Product Class 6
ssorcls1	12 Year Sales Product Class 1
ssorcls3	12 Year Sales Product Class 3
crcpr01	Circs Prom 01
crcpr17	Circs Prom 17
crcpr21	Circs Prom 21
crcpr24	Circs Prom 24
crcpr35	Circs Prom 35
crcpr50	Circs Prom 50
crcpr60	Circs Prom 60
crcpr65	Circs Prom 65

crcpr72	Circs Prom 72
crcpr85	Circs Prom 85
crcpr93	Circs Prom 93
aocls6	12 Yr Avg Ord, Product Class 6
totord	Lifetime Orders
totsale	Lifetime Sales
custaord	Lifetime Average Order
maxdol	Largest Order
rec1	Recency of last Order< 7 months (Y/N)
rec2	Recency of last Order 7-12 months (Y/N)
rec3	Recency of last Order 13-18 months (Y/N)
rec4	Recency of last Order 19-24 months (Y/N)
rec5	Recency of last Order 25-30 months (Y/N)
rec6	Recency of last Order 31-36 months (Y/N)
rec7	Recency of last Order > 36 months (Y/N)
totnumtl	Total Number of All Tradelines
avebcbal	Ave Balance of All Open+Act Bank Cards
numsatbc	Num of All Bank Cards w. Satisfac Rating
numccact	Number of All Credit Card Accounts
dmawlthr	DMA wealth rating
incmindx	income index
wealthrt	wealth rating
perperhh	persons per HH
prcncd1	% NCDB HH-1 unit structures
medschyr	median years school for people age 25+
oohvi	OOH home value index
hhmedage	householders' median age
prchhfm	% HH that are families
populat	population

Appendix 3. Model Coefficient Estimation: Original Ratio Training Data

Variables	BBR	LR	HBL2		LCA2		HBL3			LCA3		
			Class1	Class2	Class1	Class2	Class1	Class2	Class3	Class1	Class2	Class3
ord165	-0.089	-0.073	-0.044	-0.044	0.272	-46.117	0.413	0.411	0.412	29.852	-1.543	-37.951
ord172	-0.043	-0.101	-0.035	-0.035	0.583	-30.070	0.001	0.002	0.001	-47.564	11.463	-73.113
ord185	0.541	0.505	0.157	0.156	2.035	7.373	0.553	0.550	0.552	25.568	26.332	-2.813
ord193	-0.357	-0.479	-0.056	-0.056	1.175	-155.855	-0.269	-0.269	-0.269	-17.603	26.028	-17.291
ord201	0.201	0.124	0.160	0.159	0.107	62.059	-0.249	-0.245	-0.247	-22.778	14.390	-3.862
ord250	-0.424	-0.443	-0.341	-0.339	0.178	-183.760	-0.418	-0.417	-0.417	-23.725	17.026	-41.656
ord285	0.749	0.635	0.412	0.411	-1.223	165.671	1.448	1.440	1.443	35.422	-2.233	-4.830
ord301	0.501	0.550	0.201	0.202	0.295	55.111	0.461	0.461	0.460	-4.196	-32.126	7.872
ord317	0.857	0.929	0.433	0.430	3.519	-332.016	0.603	0.602	0.602	-1.593	34.116	28.657
ord350	-0.453	-0.508	-0.263	-0.261	0.064	-97.833	-0.525	-0.524	-0.525	-4.855	10.901	-13.350
ord385	1.673	1.662	0.970	0.963	2.432	48.796	1.641	1.637	1.638	19.980	17.170	21.914
ord424	-0.158	-0.165	-0.146	-0.143	-1.402	33.739	-0.659	-0.657	-0.657	-22.220	-19.716	8.625
ord435	-0.378	-0.444	-0.213	-0.211	0.062	-72.820	-0.260	-0.260	-0.260	-1.594	0.508	-8.625
ord450	-0.516	-0.594	-0.236	-0.236	0.931	-238.667	-0.513	-0.512	-0.512	-4.603	18.202	-19.325
ord485	0.306	0.269	0.024	0.025	-2.221	28.731	0.077	0.076	0.077	12.736	-11.948	10.491
ord493	-1.169	-1.396	-0.366	-0.364	-5.225	-73.028	-0.599	-0.599	-0.599	-5.522	-58.287	-24.553
fstycls3	-0.007	-0.011	-0.023	-0.004	-0.032	0.350	0.022	0.016	0.019	0.338	-0.997	-0.768
fstcls2	-0.158	-0.177	-0.087	-0.085	-1.266	16.763	0.369	0.369	0.369	14.983	-17.749	0.811
lstycls3	0.013	0.017	-0.003	-0.005	0.089	0.492	-0.001	0.074	0.042	0.078	1.016	0.363
yrord1	-0.220	-0.318	-0.064	-0.067	-0.657	20.995	0.399	0.399	0.399	4.291	-7.375	-16.714
yrord2	-0.156	-0.270	0.001	0.001	0.312	-65.130	0.290	0.287	0.288	-0.541	9.131	-7.420
yrord3	-0.256	-0.302	-0.055	-0.056	-0.397	-36.193	-0.108	-0.111	-0.110	0.623	-4.067	-9.777
yrord4	0.088	0.088	0.103	0.102	-0.042	2.072	-0.026	-0.024	-0.025	-5.781	7.454	6.956
yrord5	0.288	0.247	0.237	0.239	-0.637	36.467	0.392	0.392	0.392	4.511	5.327	5.607
yrord6	-0.351	-0.368	-0.237	-0.236	0.185	-36.108	-0.255	-0.254	-0.255	4.338	-8.511	-10.737
yrord7	-0.109	-0.128	-0.027	-0.028	-0.003	7.727	-0.355	-0.353	-0.354	-11.795	0.877	2.724
yrord8	0.563	0.559	0.421	0.419	-0.135	67.947	0.203	0.203	0.203	0.028	2.216	17.324
yrord9	-0.132	-0.161	-0.003	-0.002	1.068	-33.238	0.041	0.041	0.041	3.540	9.143	-9.965
yrord10	-0.541	-0.569	-0.363	-0.361	-0.628	-29.959	-0.174	-0.174	-0.174	-5.221	-8.971	-15.714
yrord11	0.805	0.772	0.581	0.583	0.671	14.507	0.418	0.420	0.419	-4.711	14.196	17.789
yrord12	-0.254	-0.260	-0.098	-0.098	0.069	-43.693	-0.246	-0.245	-0.246	-13.347	-11.286	8.194
prord12	-0.054	-0.012	-0.131	-0.132	0.495	-86.525	-0.309	-0.308	-0.309	-5.868	2.392	-12.724
prord80	0.043	0.033	0.033	0.032	0.241	-12.445	0.113	0.117	0.115	-3.188	9.171	-3.163
prord85	0.044	0.057	0.216	0.218	0.325	15.124	0.394	0.381	0.387	-1.889	3.405	2.055
cat18	-0.637	-0.688	-0.386	-0.384	-1.895	-9.371	-0.794	-0.792	-0.792	-53.396	20.139	-20.799
cat25	0.345	0.404	0.295	0.294	0.392	37.702	0.865	0.859	0.862	13.799	-0.506	9.868
cat26	0.137	0.135	0.143	0.142	1.527	-36.836	0.084	0.082	0.083	-6.144	16.896	-15.009

cat31	-0.407	-0.464	-0.293	-0.293	-0.834	-6.031	-0.694	-0.693	-0.693	-14.679	-6.767	-5.304
cat33	-0.190	-0.220	-0.140	-0.144	-0.364	-24.666	-0.562	-0.559	-0.560	-13.962	-7.214	1.146
cnvcat22	-0.300	-0.354	-0.051	-0.051	-0.122	-87.848	-0.206	-0.205	-0.205	-6.911	9.073	-79.098
cnvcat25	0.316	0.331	0.232	0.229	0.142	1.770	0.785	0.780	0.782	9.719	5.207	-13.196
cnvcat26	-0.091	-0.079	-0.100	-0.102	-0.549	3.880	-0.299	-0.299	-0.299	-4.598	-44.611	11.505
cnvcat33	0.022	-0.003	-0.027	-0.028	0.215	-14.937	-0.321	-0.318	-0.319	-3.140	-3.500	-0.616
fphone	0.116	0.128	0.053	0.050	0.112	2.350	-0.084	-0.080	-0.082	-9.431	14.179	-6.749
convprom	0.004	0.005	0.001	0.007	0.013	0.392	0.012	0.065	0.039	0.306	0.306	-0.128
salcat	0.154	0.161	0.168	0.164	0.303	47.080	0.254	0.254	0.254	2.944	-3.288	6.443
salflg	-0.036	-0.012	-0.027	-0.027	0.099	-30.718	0.488	0.486	0.487	5.271	0.071	-5.327
hcrd	-0.204	-0.160	-0.126	-0.125	-0.486	-18.346	-0.074	-0.075	-0.075	3.095	-0.016	-15.610
cash	-0.310	-0.288	-0.305	-0.301	-1.076	1.716	-0.303	-0.300	-0.301	-9.941	-16.153	-6.301
tele	0.808	0.948	0.315	0.314	0.076	111.845	0.525	0.524	0.524	-0.198	-13.267	18.175
bus	0.037	0.046	0.044	0.044	-0.675	4.750	0.247	0.247	0.247	0.562	-4.170	-11.465
recmon	-0.102	-0.069	-0.016	-0.043	-0.240	0.027	0.035	0.046	0.043	0.535	-0.151	-0.382
totsal1	0.002	0.002	0.000	-0.005	0.004	-1.007	-0.008	0.022	0.011	-0.011	0.093	0.156
totsal2	-0.001	-0.001	-0.003	0.000	-0.007	0.596	0.003	-0.009	0.001	0.159	-0.128	-0.016
totsal3	-0.001	-0.001	-0.003	0.007	-0.007	0.739	0.003	-0.032	-0.014	0.115	-0.157	0.127
totsal4	-0.004	-0.004	-0.006	-0.012	0.000	-0.075	-0.002	0.004	0.001	0.118	-0.143	-0.213
totsal5	-0.002	-0.001	-0.003	-0.002	0.004	0.112	0.002	0.006	0.004	0.084	-0.090	-0.012
totsal6	0.011	0.011	0.007	0.020	0.002	0.977	0.012	0.018	0.015	0.022	0.238	0.233
totsal7	0.001	0.001	0.000	-0.005	0.007	-0.753	-0.012	0.002	-0.005	0.033	0.079	0.053
totsal8	-0.015	-0.016	-0.014	-0.028	-0.020	-1.330	0.001	-0.010	-0.004	0.124	-0.357	-0.590
totsal9	0.002	0.002	-0.003	-0.012	-0.014	0.649	-0.027	-0.016	-0.022	-0.133	0.011	0.371
totsal10	0.015	0.015	0.011	0.002	0.009	0.751	0.011	0.013	0.012	0.116	0.034	0.373
totsal11	-0.022	-0.021	-0.018	0.002	-0.027	-0.084	-0.002	0.023	0.010	0.250	-0.217	-0.557
totsal12	-0.003	-0.004	-0.007	-0.025	0.007	1.002	-0.039	-0.016	-0.027	-0.160	0.089	-0.197
ordcls2	0.336	0.439	0.164	0.160	0.538	28.670	-0.260	-0.255	-0.257	-3.231	7.547	5.963
ordcls7	0.276	0.346	0.090	0.091	0.244	-1.061	0.297	0.299	0.298	-1.694	3.509	10.330
ordcls3	0.665	0.789	0.405	0.403	0.735	-5.182	0.653	0.643	0.648	2.830	0.547	13.508
sordcls6	-0.057	-0.043	-0.093	-0.085	-0.311	-4.801	0.187	0.190	0.189	1.817	-2.668	-1.786
ssorcls5	0.000	0.000	0.000	-0.001	-0.001	-0.036	-0.001	0.002	0.001	-0.001	-0.010	-0.007
ssorcls6	0.004	0.004	0.003	0.013	0.006	1.035	-0.001	0.034	0.017	0.013	0.065	0.190
ssorcls1	0.001	0.001	0.000	-0.010	0.004	0.181	0.005	0.043	0.024	0.002	0.024	0.087
ssorcls3	0.000	0.000	0.000	0.000	0.005	0.307	0.003	-0.013	-0.004	-0.001	0.051	0.053
crcpr01	-0.047	-0.038	-0.094	-0.092	-0.027	-0.360	0.068	0.069	0.069	1.859	-0.578	-1.861
crcpr17	-0.001	0.014	-0.012	-0.012	0.384	-35.282	-0.300	-0.295	-0.297	-5.645	1.519	-6.038
crcpr21	-1.082	-1.054	-0.794	-0.790	-1.446	-7.847	-0.996	-0.999	-0.997	9.405	-22.799	7.377
crcpr24	0.198	0.180	0.173	0.170	0.234	25.049	0.206	0.206	0.206	0.305	1.459	5.889
crcpr35	-0.097	-0.084	-0.126	-0.126	-0.137	2.384	0.252	0.258	0.255	-2.024	13.998	-12.805
crcpr50	0.604	0.634	0.531	0.522	0.415	45.568	1.159	1.141	1.150	11.603	1.418	7.874
crcpr60	0.379	0.363	0.294	0.294	-0.065	43.329	0.437	0.440	0.438	5.740	-4.368	10.541
crcpr65	-0.245	-0.231	-0.252	-0.252	-0.292	-8.641	-0.933	-0.936	-0.934	-7.803	-4.765	-2.302

crcpr72	-0.291	-0.308	-0.225	-0.222	0.332	-33.166	-0.180	-0.178	-0.179	-0.775	3.504	-7.287
crcpr85	-0.004	0.045	-0.010	-0.015	0.205	-3.274	-0.241	-0.256	-0.248	-3.130	-0.527	3.209
crcpr93	0.001	0.021	0.023	0.031	-0.098	-11.152	0.342	0.342	0.342	2.651	3.915	0.720
aocls6	-0.005	-0.004	-0.003	-0.055	-0.005	-1.575	-0.034	0.048	0.008	-0.501	0.014	-0.296
totord	0.024	0.039	-0.002	0.012	0.090	1.202	-0.037	-0.035	-0.036	2.229	-0.824	0.223
totsale	-0.001	0.000	0.001	0.002	0.000	-0.313	0.000	0.012	0.014	-0.051	0.012	-0.046
custaord	-0.002	-0.003	-0.004	0.005	0.003	-1.404	-0.023	0.017	0.004	-0.317	0.081	-0.180
maxdol	0.004	0.003	0.004	0.002	0.001	-0.430	-0.003	-0.016	-0.002	-0.072	0.110	-0.054
rec1	-3.132	-3.680	-0.466	-0.465	-8.887	-85.870	-1.434	-1.432	-1.432	-42.570	-5.667	-5.888
rec2	-2.821	-3.539	-0.485	-0.482	-8.346	-72.320	-1.471	-1.472	-1.471	-34.792	-13.713	3.067
rec3	-2.203	-3.190	-0.494	-0.488	-7.313	-85.270	-1.326	-1.325	-1.325	-23.616	-10.962	-7.735
rec4	-1.597	-2.711	-0.350	-0.348	-5.411	-102.447	-1.155	-1.153	-1.153	-32.360	-2.141	-4.610
rec5	-0.086	-1.382	-0.204	-0.203	-5.185	18.434	-0.470	-0.468	-0.469	-1.067	-15.270	-12.188
rec6	-0.214	-1.232	-0.202	-0.200	-10.832	-8.879	-0.624	-0.624	-0.623	-9.918	2.242	-17.521
rec7	1.764	0.000	0.385	0.383	-0.303	0.087	0.794	0.794	0.793	0.052	-0.147	-0.213
totnumtl	-0.020	0.012	-0.130	-0.124	0.068	3.805	-0.168	-0.165	-0.164	-2.650	-0.108	4.392
avebcbal	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000	0.001	0.002	-0.001	-0.005
numsatbc	0.130	0.180	0.111	0.110	0.401	1.794	-0.479	-0.477	-0.477	-5.054	1.410	6.220
numccact	0.050	0.191	0.121	0.120	0.921	-67.096	0.428	0.426	0.427	22.606	-6.410	-23.999
dmawlthr	-0.025	-0.019	-0.003	-0.002	-0.151	6.841	0.054	0.044	0.050	1.792	-3.865	1.620
incmindx	-0.001	0.002	-0.004	-0.003	0.001	0.200	0.020	0.036	0.048	-0.014	0.091	-0.008
wealthrt	-0.034	-0.070	-0.014	-0.017	0.057	-8.904	-0.150	-0.153	-0.151	-0.709	3.086	-4.917
perperhh	0.234	0.294	-0.033	-0.031	0.051	-1.823	0.367	0.365	0.366	7.889	-0.159	0.114
prcncd1	-0.007	-0.008	-0.012	-0.001	-0.009	-0.063	-0.011	-0.041	-0.008	0.363	-0.281	-0.005
medschyr	-0.027	-0.042	-0.066	-0.056	-0.591	13.080	-0.216	-0.216	-0.213	4.086	-0.332	-2.582
oohvi	0.002	0.002	0.004	0.001	0.009	-0.176	-0.001	-0.008	0.015	-0.013	-0.051	0.115
hhmedage	-0.003	0.002	-0.019	-0.020	0.012	0.907	-0.100	-0.113	-0.094	-0.150	-0.453	0.310
prchhfm	0.009	0.011	0.021	0.017	0.020	0.128	0.021	0.040	0.046	-0.135	0.610	-0.284
populat	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
constant	-0.750	-1.538	-0.504	-0.502	0.302	-0.302	-0.673	-0.672	-0.672	0.700	0.007	-0.708

Appendix 4. Model Coefficient Estimation: Balanced Ratio Training Data

Variables	BBR	LR	HBL2		LCA2		HBL3			LCA3		
			Class1	Class2	Class1	Class2	Class1	Class2	Class3	Class1	Class2	Class3
ord165	0.061	2.029	-0.054	-0.053	-7.037	24.873	0.440	0.001	0.001	14.472	-13.273	19.477
ord172	-0.036	0.496	-0.047	-0.046	2.781	-4.263	-0.446	-0.001	-0.001	-7.623	5.058	1.128
ord185	-0.072	-2.476	-0.138	-0.140	-6.203	-6.855	-0.117	-0.027	-0.001	6.906	-18.043	-9.063
ord193	-0.004	-0.765	0.138	0.137	15.515	-8.023	-0.261	0.007	0.000	-2.089	-12.824	9.086
ord201	0.075	0.225	0.101	0.098	9.890	5.035	0.292	-0.008	0.000	1.223	10.081	2.974
ord250	-0.125	-0.511	-0.255	-0.254	-2.761	-3.313	-0.981	0.006	-0.002	-3.764	7.504	-5.225
ord285	0.189	-0.732	0.357	0.355	12.933	0.946	0.568	0.010	0.001	11.343	5.719	-6.018
ord301	-0.119	-0.037	-0.230	-0.226	-14.789	0.670	-0.471	0.003	-0.001	-1.972	10.674	-8.216
ord317	0.114	5.615	0.319	0.318	21.383	31.414	0.810	0.001	0.001	19.240	35.836	39.433
ord350	-0.029	-1.019	-0.229	-0.227	-16.392	-21.757	-0.107	-0.001	0.000	3.916	-6.475	-7.936
ord385	0.258	2.500	0.609	0.606	10.949	14.609	1.490	0.002	0.002	17.762	34.663	10.404
ord424	0.058	1.141	0.143	0.143	5.572	-4.093	0.308	-0.002	0.001	4.628	11.358	5.677
ord435	0.049	1.773	0.270	0.268	15.620	11.695	0.176	0.000	0.000	13.876	14.943	3.288
ord450	-0.013	0.082	0.168	0.166	9.505	-6.250	-0.005	0.002	0.000	-7.368	9.133	6.243
ord485	-0.026	-0.976	0.007	0.008	3.994	-0.459	-0.255	0.003	0.000	-20.972	-3.705	4.173
ord493	0.036	-0.556	0.056	0.055	-19.801	-1.939	0.324	0.001	0.001	5.011	-6.125	-2.854
fstycls3	-0.037	-0.120	-0.021	-0.008	-0.201	-0.240	-0.090	0.031	-0.002	-0.563	-0.586	0.040
fstcls2	0.002	-0.209	0.075	0.073	3.827	-7.308	-0.031	0.015	0.000	-1.956	-0.911	4.048
lstycls3	-0.006	0.025	-0.014	-0.007	-0.007	0.408	-0.032	0.176	0.018	0.074	0.043	0.393
yrord1	-0.199	-1.120	-0.586	-0.585	-17.266	-14.439	-0.342	0.002	-0.001	-12.796	-9.576	-5.606
yrord2	0.032	0.076	-0.059	-0.060	-10.136	1.007	-0.227	0.005	0.000	-2.579	-2.112	3.948
yrord3	0.138	0.099	0.158	0.162	-3.361	8.007	0.169	0.002	0.000	-2.603	-3.163	3.700
yrord4	0.099	1.054	0.237	0.238	5.479	-4.706	0.246	0.003	0.001	8.029	-0.278	4.534
yrord5	0.099	0.818	0.306	0.305	5.231	1.613	0.410	0.004	0.001	12.143	6.634	1.433
yrord6	-0.095	-0.286	-0.243	-0.242	0.481	-10.408	-0.336	-0.001	-0.001	0.117	-11.917	-0.596
yrord7	-0.078	-0.011	-0.189	-0.189	0.659	0.955	-0.348	-0.002	0.000	-6.788	-1.113	5.719
yrord8	0.213	1.691	0.716	0.712	11.119	8.682	0.380	0.002	0.002	3.121	7.882	3.317
yrord9	-0.028	-0.589	0.118	0.117	-2.881	-7.318	-0.045	-0.002	0.000	1.930	8.135	-12.777
yrord10	-0.251	-2.787	-0.588	-0.585	-14.432	-14.499	-0.103	-0.003	-0.001	-7.435	-17.933	-9.513
yrord11	0.136	2.508	0.470	0.468	4.399	15.158	0.313	0.001	0.001	15.386	14.354	14.391
yrord12	-0.003	2.010	0.083	0.084	1.282	8.224	0.090	-0.002	0.001	9.610	2.908	10.954
prord12	-0.080	0.460	-0.418	-0.416	-10.323	1.433	-0.415	-0.003	-0.001	3.523	-11.833	-3.991
prord80	-0.068	-0.013	0.024	0.023	2.959	-3.615	-0.075	-0.002	0.000	-2.565	-2.976	-4.476
prord85	0.260	1.314	0.498	0.494	0.356	9.760	0.274	-0.013	0.000	10.779	5.785	5.276
cat18	-0.173	-3.002	-0.420	-0.419	-14.504	-17.367	-0.985	-0.002	-0.002	-15.614	-10.946	-20.267
cat25	0.144	1.970	0.312	0.311	5.103	7.594	0.290	0.002	0.000	3.877	1.937	10.572
cat26	0.006	-0.442	-0.078	-0.079	-5.114	4.006	-0.196	0.006	0.000	-2.110	2.431	1.076

cat31	-0.152	-1.583	-0.276	-0.274	-4.925	0.394	-0.636	0.007	-0.001	-6.225	-3.677	-1.147
cat33	0.022	-0.500	0.010	0.009	1.844	2.154	0.415	-0.004	0.001	6.158	2.757	-1.172
cnvcat22	-0.020	-0.127	-0.159	-0.158	-5.847	9.571	0.525	0.000	0.001	17.867	-9.797	-2.889
cnvcat25	0.060	0.168	0.043	0.043	0.951	8.235	0.006	0.001	0.000	0.088	7.299	0.774
cnvcat26	-0.032	-0.808	-0.074	-0.075	-4.777	-7.316	-0.052	-0.027	0.000	-2.873	-0.010	4.309
cnvcat33	-0.079	-0.696	-0.277	-0.274	-15.278	1.932	-0.339	-0.005	-0.001	-2.704	1.974	-9.646
fphone	0.023	-0.339	0.142	0.137	6.973	-2.733	-0.298	-0.006	0.000	-1.928	1.908	1.383
convprom	0.002	0.012	0.005	-0.010	0.050	0.134	0.002	-0.053	-0.011	-0.026	-0.021	0.205
salcat	-0.060	-0.161	0.121	0.115	5.919	-4.194	-0.071	-0.020	-0.002	-1.296	-0.538	-1.465
salflg	-0.080	0.028	0.045	0.038	-1.080	-1.380	-0.022	-0.079	-0.002	1.102	-1.364	-2.405
hcrd	0.050	-0.199	0.263	0.262	1.036	-8.202	-0.137	0.004	0.000	2.321	-4.564	-4.527
cash	-0.173	-0.969	-0.228	-0.224	-4.000	-9.551	-0.278	0.003	-0.001	-0.542	-1.529	-9.172
tele	0.058	2.137	0.236	0.235	9.127	6.982	0.203	0.005	0.001	9.786	10.929	8.788
bus	-0.035	-0.104	-0.047	-0.047	-0.716	-0.153	-0.080	0.000	0.000	-0.425	-1.451	-0.258
recmon	-0.018	-0.226	-0.028	0.041	-0.533	-0.819	-0.058	-0.059	0.017	-0.806	-1.320	-0.936
totsal1	0.002	-0.013	-0.001	0.005	0.015	0.114	0.000	0.011	-0.034	-0.036	0.014	0.055
totsal2	-0.005	-0.023	-0.009	0.007	-0.099	0.017	0.001	-0.034	0.011	-0.065	-0.076	-0.101
totsal3	0.000	-0.017	-0.005	0.004	0.012	0.000	-0.002	0.081	-0.041	-0.028	0.006	-0.006
totsal4	-0.003	-0.037	-0.003	-0.026	-0.089	0.016	-0.002	0.030	-0.009	-0.141	-0.003	-0.214
totsal5	0.001	-0.016	-0.005	0.023	-0.199	-0.014	-0.006	0.065	-0.001	-0.189	-0.118	-0.124
totsal6	0.011	-0.005	0.012	-0.018	0.083	0.355	0.030	-0.038	0.000	0.129	0.259	0.109
totsal7	-0.003	-0.018	-0.004	-0.040	-0.103	-0.207	-0.009	0.003	0.020	-0.106	0.002	-0.131
totsal8	-0.012	-0.065	-0.024	-0.038	-0.371	-0.384	-0.027	-0.013	0.019	-0.200	-0.206	-0.097
totsal9	0.001	-0.007	0.000	-0.044	0.075	0.166	0.011	-0.047	0.007	0.022	-0.320	0.222
totsal10	0.012	0.060	0.010	0.009	0.187	0.587	0.023	-0.070	-0.006	0.259	0.565	0.294
totsal11	-0.003	-0.078	-0.019	-0.014	-0.123	-0.267	-0.021	0.045	0.000	-0.569	-0.079	-0.294
totsal12	0.012	-0.088	-0.016	0.023	-0.370	-0.205	0.024	-0.073	0.012	-0.203	-0.562	-0.292
ordcls2	0.123	1.229	0.129	0.128	4.940	8.740	0.223	0.015	0.000	9.756	11.553	8.606
ordcls7	0.002	1.183	0.217	0.215	17.197	4.329	0.227	0.008	0.000	13.218	6.011	-1.897
ordcls3	0.201	3.020	0.425	0.420	20.069	9.477	0.365	-0.019	0.000	-1.066	17.743	8.581
sordcls6	-0.256	-0.684	-0.326	-0.323	-2.714	-7.824	-0.127	0.000	0.001	-6.008	-3.072	-4.016
ssorcls5	0.000	-0.001	-0.001	-0.022	-0.007	-0.007	0.000	0.006	0.018	-0.006	-0.006	-0.011
ssorcls6	0.006	0.020	0.012	0.011	0.149	0.229	0.011	-0.033	0.008	0.223	0.110	0.083
ssorcls1	-0.001	-0.001	0.000	-0.018	0.012	0.008	-0.002	-0.046	0.014	0.028	-0.007	-0.021
ssorcls3	0.003	-0.001	0.004	-0.065	0.078	0.022	0.006	-0.078	-0.076	0.018	0.021	0.079
crcpr01	-0.122	-0.078	-0.178	-0.179	1.621	-3.076	-0.141	0.088	-0.001	-3.200	-0.808	-1.084
crcpr17	-0.163	-0.615	-0.367	-0.362	-10.200	-0.912	-0.217	0.014	0.000	-6.043	-1.924	-5.985
crcpr21	-0.297	-2.106	-0.795	-0.793	-10.375	-22.854	-0.630	0.000	-0.001	-10.593	-9.811	-11.443
crcpr24	-0.018	-0.222	-0.281	-0.274	-7.565	1.655	-0.189	0.102	0.001	-1.417	-1.114	1.670
crcpr35	-0.087	-0.752	-0.628	-0.618	-19.422	6.170	-1.058	0.010	-0.001	-14.176	1.067	3.890
crcpr50	0.503	1.199	0.898	0.880	13.869	5.042	0.560	-0.055	-0.001	8.209	10.754	6.622
crcpr60	0.140	1.466	0.593	0.590	13.525	9.111	0.401	-0.003	0.001	4.579	4.541	8.661
crcpr65	-0.078	0.239	-0.118	-0.122	-1.625	-0.205	-0.041	-0.023	0.001	1.676	-8.461	-3.204

crcpr72	-0.175	-0.587	-0.163	-0.163	-2.686	-0.108	-0.419	0.021	0.000	-3.877	-7.945	-6.002
crcpr85	0.119	0.235	0.148	0.140	0.897	1.493	0.199	0.074	-0.002	3.243	2.568	1.934
crcpr93	0.017	-0.325	-0.180	-0.182	-6.647	1.095	0.270	-0.011	0.001	0.250	-3.103	-2.395
aocls6	-0.005	-0.025	-0.016	0.019	-0.151	-0.094	-0.004	0.036	-0.008	-0.144	-0.064	-0.100
totord	0.103	-0.149	0.033	0.040	0.304	3.312	0.019	0.143	0.001	-0.389	0.282	0.777
totsale	-0.001	0.016	0.001	0.016	0.032	-0.103	-0.002	0.019	-0.038	0.038	0.011	-0.004
custaord	0.004	0.005	-0.010	0.017	-0.270	0.157	0.015	-0.058	-0.036	0.174	-0.011	-0.134
maxdol	0.002	0.001	0.009	-0.034	0.062	0.029	0.004	-0.036	-0.038	-0.045	-0.037	0.117
rec1	-0.093	-10.64	-0.295	-0.293	-23.460	-38.406	-0.061	-0.030	0.000	-29.339	-37.490	-39.616
rec2	-0.063	-10.43	-0.191	-0.191	-30.058	-41.400	-0.104	-0.026	-0.001	-31.546	-55.750	-40.129
rec3	-0.151	-10.84	-0.200	-0.201	-24.831	-75.041	-0.135	-0.017	0.000	-26.466	-49.747	-48.116
rec4	-0.085	-9.691	-0.510	-0.507	-38.431	-42.747	-0.805	0.008	-0.001	-36.969	-42.087	-35.705
rec5	-0.021	-5.612	-0.059	-0.061	-20.489	-27.857	-0.246	0.000	0.000	-28.971	-14.353	-16.516
rec6	-0.046	-5.749	-0.075	-0.074	-19.684	-45.588	-0.007	0.000	0.000	-22.370	-36.806	-29.220
rec7	0.097	0.000	0.354	0.354	-0.158	0.198	0.427	0.006	0.001	0.049	0.148	0.164
totnumtl	0.020	0.351	0.101	0.096	1.992	1.774	-0.043	0.183	0.003	-0.631	4.042	2.361
avebcbal	0.000	-0.001	-0.001	-0.001	-0.007	-0.009	-0.001	0.001	0.002	-0.009	-0.008	-0.009
numsatbc	0.220	0.695	0.263	0.262	5.191	2.523	0.211	0.006	0.001	7.516	1.523	3.109
numccact	0.124	2.434	0.389	0.388	16.196	18.470	0.144	-0.003	0.000	16.369	11.816	23.744
dmawlthr	-0.031	0.003	-0.045	-0.038	1.392	0.193	-0.269	0.057	0.002	-2.477	-0.151	0.605
incmindx	-0.002	-0.012	-0.011	0.026	-0.045	-0.190	-0.014	0.032	0.024	-0.045	-0.122	-0.147
wealthrt	-0.065	-0.328	-0.029	-0.029	-3.114	1.252	0.232	-0.006	0.001	1.100	-1.738	0.839
perperhh	-0.123	-0.252	-0.257	-0.256	-7.569	1.443	0.017	0.000	0.000	-8.292	7.696	-3.393
prcncd1	0.004	0.014	0.013	-0.011	0.111	0.220	-0.013	-0.005	0.023	-0.269	0.374	0.273
medschyr	-0.038	0.084	-0.086	-0.087	-6.555	2.660	0.068	-0.008	0.003	-4.073	3.689	2.950
oohvi	0.005	0.016	0.007	0.020	0.090	0.136	0.003	0.007	0.040	0.071	0.071	0.002
hhmedage	-0.011	0.037	0.001	-0.032	0.524	-0.148	0.010	-0.010	0.014	-0.369	0.343	0.750
prchhfm	0.005	0.020	0.012	-0.024	0.047	-0.153	0.032	-0.072	0.011	0.640	-0.337	-0.008
populat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
constant	-0.044	3.958	-0.098	-0.098	0.044	-0.044	-0.307	0.001	0.000	0.140	-0.064	-0.076

References

Allenby, G. M., Bakken, D. G., and Rossi, P. E. (2004). "The HB Revolution: How Bayesian Methods Have Changed the Face of Marketing Research". *Marketing Research*, 20-25.

Allenby, G. M., Leone, R. P., and Jen, L. (1999). "A Dynamic Model of Purchase Timing with Application to Direct Marketing". *Journal of the American Statistical Association*, Vol. 94, No. 446, 365-374.

Andrews, R. L., and Currim, I. S. (2003). "A Comparison of Segment Retention Criteria for Finite Mixture Logit Models". *Journal of Marketing Research*, Vol. 40, No. 2, 235-243.

Baesens, B., Viaene, S., Poel, D. V., Vanthienen, J., and Dedene, G. (2002). "Bayesian Neural Network Learning for Repeat Purchase Modeling in Direct Marketing". *European Journal of Operational Research*, Vol. 138, 191-211.

Berger, P., and Magliozzi, T. (1992). "The Effect of Sample Size and Proportion of Buyers in the Sample on the Performance of List Segmentation Equations Generated by Regression Analysis". *Journal of Direct Marketing*, Vol. 6, No. 1, 13-22.

Bitran, G. R., and Mondschein, S. V. (1996). "Mailing Decisions in the Catalog Sales Industry". *Management Science*, Vol.42, No. 9, 1364-1381.

Bult, J. R., and Wansbeek, T. (1995). "Optimal Selection for Direct Mail". *Marketing Science*, Vol. 14, No. 4, 378-394.

Colombo, R., and Jiang, W. (1999). "A Stochastic RFM Model". *Journal of Interactive Marketing*, Vol. 13, No. 3, 2-12.

Cui, G., and Wong, M. L. (2004). "Implementing Neural Networks for Decision Support in Direct Marketing". *International Journal of Market Research*, Vol. 46, 235-254.

Cui, G., Wong, M. L., and Lui H. K. (2006). "Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming". *Management Science*, Vol. 52, No. 4, 597-612.

Direct Marketing Association (2006). "Executive Summary for 2005" and "The DMA 2005 Response Rate Report". *http://www.the-dma.org/*.

Evans, M., Nairn, A., and Maltby, A. (2000). "The Hidden Sex Life of the Male (and Female) Shot". *International Journal of Advertising*, Vol. 19, 1, 43-65.

Fawcett, T. (2006). "An Introduction to ROC Analysis". *Pattern Recognition Letters*, Vol. 27, 861-874.

Genkin, A., Lewis, D. D., and Madigan, D. (2005) "Large-Scale Bayesian Logistic Regression for Text Categorization". *Technique Report*, *DIMACS*, 1-43.

Gonul, F. F., and Hofstede, F. T. (2006). "How to Compute Optimal Catalog Mailing Decisions". *Marketing Science*, Vol. 25, No. 1, 65-74.

Haughton, D., and Oulabi, S. (1997). "Direct Marketing Modeling with CART and CHAID". *Journal of Direct Marketing*, Vol. 11, No. 4, 42-52.

Heinen, T. (1996). "Latent Class and Discrete Latent Trait Models: Similarities and Differences". *Sage Publications*.

Ho, Y. C., and Pepyne, D. L. "Simple Explanation of the No-Free-Lunch Theorem and Its Implications". *Journal of Optimization Theory and Applications*, Vol. 115, No. 3, 549-570.

Jain, D., Bass, F. M., and Chen, Y. M. (1990). “Estimation of Latent Class Models with Heterogeneous Choice Probabilities: An Application to Market Structuring”. *Journal of Marketing Research*, Vol. 27, No. 1, 94-101.

Jedidi, K., Jagpal, H. S., and Desarbo, W. S. (1997). “Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity”. *Marketing Science*, Vol. 16, No. 1, 39-59.

Jen, L., Chou, C., and Allenby, G. M. (2003). “A Bayesian Approach to Modeling Purchase Frequency”. *Marketing Letters*, Vol 14, No. 1, 5-20.

Langseth, H., and Nielsen, T. D. (2006). “Classification using Hierarchical Naïve Bayes Models”. *Machine Learning*, Vol. 63, 135-159.

Linder, R., Geier, J., and Kolliker, M. (2004). “Artificial Neural Networks, Classification Trees and Regression: Which Method for Which Customer”. *Journal of Database Marketing and Customer Strategy Management*, Vol. 11, No. 4, 344-356.

Natter, M., and Feurstein, M. (2002). “Real World Performance of Choice-Based Conjoint Models”. *European Journal of Operational Research*, Vol. 137, 448-458.

Neelamegham, R., and Chintagunta, P. (1999). “A Bayesian Model to Forecast New Product Performance in Domestic and International Markets”. *Marketing Science*, Vol. 18, No. 2, 115-136.

Robert, C., and Casella, G. (2004). “Monte Carlo Statistical Methods”. *Springer Verlag*.

Rossi, P. E., McCulloch, R. E., and Allenby, G. M. (1996). “The Value of Purchase History Data in Target Marketing”. *Marketing Science*, Vol. 15, No. 4, 321-340.

Rossi, P. E., and Allenby, G. M. (2003). “Bayesian Statistics and Marketing”, *Marketing Science*, 2003, Vol.22, No.3, 304-328.

Rossi, P. E., Allenby, G. M., and McCulloch, R. (2006). “Bayesian Statistics and Marketing”, *Wiley Series in Probability and Statistics*.

Rud, O. P. (2001). “Data Mining Cookbook: Modeling Data for Marketing, Risk and Customer Relationship Management”. *Wiley*.

Shepard, D. (1999). “The New Direct Marketing: How to Implement a Profit-Driven Database Marketing Strategy”. *McGraw-Hill*.

Steenburgh, T. J., Ainslie, A., and Engebretson, P. H. (2003). “Massively Categorical Variables: Revealing the Information in Zip Codes”. *Marketing Science*, Vol. 22, No. 1, 40-57.

Thrasher, R. P. (1991). “CART: A Recent Advance in Tree-Structured List Segmentation Methodology”. *Journal of Direct Marketing*, Vol. 5, No. 1, 35-47.

Venugopal, V., and Baets, W. (1994). “Neural Networks and their Applications in Marketing Management”. *Journal of Systems Management*, Vol. 45, No. 9, 16-21.

Wedel, M., Kamakura, W., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S., Poulsen, C. T. (1999). “Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modeling”. *Marketing Letters*, Vol. 10, No. 3, 219-232.

Wedel, M., Desarbo, W. S., Bult, J. R., and Ramaswamy, V. (1993). “A Latent Class Poisson Regression Model for Heterogeneous Count Data”. *Journal of Applied Econometrics*, Vol. 8, No. 4, 397-411.

Zahavi, J., and Levin, N. (1997). “Issues and Problems in Applying Neural Computing to Target Marketing”. *Journal of Direct Marketing*, 1997, Vol. 11, No. 4, 63-75.

Zahavi, J., and Levin, N. (1997). “Applied Neural Computing to Target Marketing”. *Journal of*

Direct Marketing, Vol. 11, No. 4, 76-93.

Zhang, N. L. (2004). "Hierarchical Latent Class Models for Cluster Analysis". *Journal of Machine Learning Research*, Vol. 5, 697-723.