CUSTOMER RESPONSE MODELS: WHAT DATA PREDICTS BEST, HARD OR SOFT?

WILLIAM J. HAUSER, The University of Akron
LINDA ORR, The University of Akron
TERRY DAUGHERTY, The University of Akron

The concept of marketing segmentation emerged in the1950s with simple demographic and psychographic components. With the advent of the information age and intensified competition, marketers have adapted more sophisticated segmentation approaches such as level of customer profitability. While both of these segmentation approaches are widely accepted, researchers have noted the need to combine these methods to enhance the current understanding of customer response models. The goal of this research is to test such a model in order to offer marketers a more efficient and effective way to tailor their marketing campaigns and allocate resources. This study demonstrates the added value of incorporating both hard (actual customer purchase) data with the softer (demographic and psychographic) data. In order to conduct the analyses, a large database was secured of over 175,000 customer purchases over a two-year time span. Within the database, each individual consumer purchase was matched against advertising exposure as well as demographic and psychographic data available from secondary sources. The goal is to not only build an applicable customer response model, but more importantly, to examine comparative models in order to assess the relative importance and contribution of each type of data. The purpose of this analysis is to suggest predictive, more cost effective, customer response profiles for both practitioners and academics struggling to better understand how to predict consumer behavior.

INTRODUCTION

Traditionally, businesses have searched for the most expeditious and accurate ways to identify those customers who are most likely to make initial and repeat purchases of their products and services. Prior to automation these processes focused on personal one to one relationships with long term customers and their families. During the later part of the 20th century these personalized one to one relationships have shifted to more structured and, in many ways, more ambiguous groupings of individuals with similarly defined traits. In turn, the technological advances of the early 21st century have allowed marketers to blend the historical one to one relationships of the past with the macroscopic data driven segmentation schemes of the present. These groupings and segments have made the customer identification processes more expeditious, but have they accurately identified the firm's most valued and valuable customers?

The Marketing Management Journal
Volume 21, Issue 1, Pages 1-15

Even with the current technological advances, some of the most important questions businesses still face is which customers to target and how many resources to invest in each target market. While a plethora of customer information data is available, businesses need to know what data to use in order to answer a number of different strategic questions. In some cases, consumer shopping data (e.g., scanner data) have been used to attempt to model actual customer responses (i.e., hard data), while others have turned to syndicated data (i.e., soft data) in attempt to build possible segments of potentially profitable customers. In today's customer-centric environment the trend has shifted toward customer profitability and the lifetime value of a customer. Kumar (2006) indicates that evaluating the future profitability of customers is crucial for strategic planning of the firm, but feels that neither researchers nor firms fully understand this concept. He posits that businesses still do not know which metric is best in assessing the lifetime value of a customer. Can customers be evaluated based only on their past contribution to the firm? Conversely, should analysts also consider other demographic and psychographic variables?

These questions and decisions are based on a rich array of segmentation and customer profitability research. The concept of marketing segmentation gained popularity in the 1950s with simple demographic segmentation (e.g., a target market of women aged 30-50). Shortly afterwards marketers attempted to increase their knowledge of their customer base through the use of psychographic or lifestyle segmentation (e.g., Prizm). With the advent of the current information age and intense competition facing many firms today, marketers have begun using more sophisticated behavioral segmentation approaches, such as customer profitability (e.g., A, B, and C accounts) and customer lifetime value (CLV). Incorporating customer profitability into segmentation strategies has enabled businesses to improve the quality of their customer knowledge base and their marketing decisions (Niraj, Gupta and Narasimhan 2001). Thus, it appears that the probability a customer will choose to buy a company's products is predicated on a wide range of variables and segmentation schemes, including demographic and psychographic data, past purchasing behavior, and the nature and quantity of the businesses' contact with the customer (Kumar, Venkatesan and Reinartz 2006).

Conceptualizing Customer Segmentation

Initially, marketing began with the simple notion of, "a good product will sell itself." The perfect example of this is Henry Ford's Model-T Ford. The car had only one model and only one color: black. In the 1950s, marketing segmentation was introduced to the discipline to reflect a change from mass marketing and the one size fits all approach to the new (at the time) marketing concept: targeting products and marketing campaigns to specific groups of customers (Lemon and Mark 2006; Tedlow 1990). Since its adoption, marketers have most frequently used demographic (i.e., age, race, or gender) and psychographic dimensions (i.e., values, attitudes, and lifestyles) to segment customers into homogenous groups (Peltier and Scribrowsky 1997). Traditional segmentation methods have received widespread attention over the past sixty years and have become a fundamental principle of the marketing discipline. These target marketing and segmenting approaches are now commonplace with any firm looking to assess and analyze who their customers are, what products and services to produce, and how to best reach these customers. While segmentation can be particularly valuable to a new company or a new product with no existing transaction data, such a specific focus can result in marketers overlooking key elements of profitability and its dynamics (Reinartz and Kumar 2000, 2002; Bolton, Lemon and Veroef 2004; Thomas, Reinartz and Kumar 2004).

Marketing experienced a significant shift in the 21st century in order to adapt to changing markets, consumer behavior patterns, and new technologies available to researchers. Gone are the days of mass marketing and so too are the days of simplistic segmentation models (Thomas, Lewison, Hauser and Foley 2007). With such saturated hypercompetitive environments, marketers simply cannot compete any longer by utilizing simple standalone factors like demographics and psychographics. Consumers are becoming much more educated and sophisticated. Thus, marketers must also become more educated in terms of how to reach these customers. Likewise, much of the demographic and psychographic data is purchased as secondary data from syndicated companies and only available at the zip code level rather than at the individual consumer level.

Profitability Modeling

While traditional segmentation may help lead to higher revenues and responsiveness, they tend to be costly to build and maintain and, most importantly, do not typically take into account profitability levels. Within the past two decades marketers began employing segmentation methods utilizing both retention rates and acquisition rates. In order to do so, analysts have turned to more behavioral types of data, such as consumer purchase data. In fact, many times this type of data is already available to the company via their internal databases. Terms like key account management and customer profitability have emerged as markets began segmenting their customers based on continuums from the most profitable to the least profitable customers. Initial models were very simplistic and required little information and statistical analyses.

The RFM method (i.e., recency, frequency, and monetary value) has been one of the most widely used methods for identifying the most profitable customers (Keiningham, Aksoy and Bejou 2006; Hughes 1996). Based on the assumption that past purchase behavior can be utilized to predict and segment a firm's most profitable customers, the premise of the model is that the most recent, most frequent, and largest spending customers are the best customers since it is assumed they will act similarly in the future as well. The belief is that these customers are also probably the most profitable ones (Keiningham, Aksoy and Bejou 2006). Many companies continue to use this basic approach to identify segments of customers, such as the A, B, and C accounts. For example, Zeithaml, Rust and Lemon (2001), suggest segmenting customers into four different tiers: platinum, gold, iron, and lead. Numerous other comparable typologies have been used such as with banks that rate their customers from a "P1" for most profitable to a "P5" for not at all profitable.

Firms have also turned toward utilizing total revenue approaches where total sales dollars from each customer are identified and analyzed in order to optimize profitability. However, correlating revenue for each customer with profitability is very risky and, in many cases, an inaccurate task. For example, "lead customers" mentioned above are those that continually cost the company more money. They demand more attention than their spending warrants and are many times big complainers, which in turn ties up too much of a firm's resources (Zeithaml, Rust and Lemon 2001).

The analytical shift toward actual customer purchase histories and the evaluation of customer profitability appears to be a step in the right direction. What a customer actually purchases is most likely a better determinant of future behavior than macroscopic variables such as age or race. However, these methods are still subject to the problem that the data models are based on past performance. What is needed is a forward-looking metric of the stream of profits (or losses) that a customer can be expected to contribute to the firm: in essence, the net present value of a customer's future stream of earnings or customer lifetime value (Keiningham, Aksoy and Bejou 2006).

Customer Lifetime Value (CLV) Segmentation

The difficulties associated with identifying valuable data, combined with the need to maximize revenue, has led to the segmentation approach referred to as Customer Lifetime Value (CLV) based segmentation. In its simplest form, CLV represents the present value of benefits less the burdens from customers (Dwyer 1989; Bolton, Lemon, Verhoef 2004). Dwyer (1989) and Blattberg and Deighton (1996) were some of the first marketing researchers to suggest the need to combine both customer acquisition and retention costs into the same analysis model. In terms of retention costs, these models incorporate marketing costs that are incurred for every customer in each period. Then, in its most basic formulation, the net present value of these earnings produces the resultant CLV (Keiningham, Aksoy and Bejou 2006; Dwyer 1989).

However, even though the premise of CLV seems simple at face value, the actual assessment of CLV is a complicated task. It is very difficult to precisely assess some of the needed variables, such as, future revenues, expenditures, and churn statistics. Nevertheless, with the rapid growth of database mining, the vast availability of data on customers, and the abundance of models available, the task of assessing CLV is becoming widely accepted (Labai, Narayandas and Humby 2002; Kumar, Venkatesan and Reinartz 2006; Lemon and Mark 2006). In short, this segmentation approach is rapidly becoming accepted as one of the most preferred dimensions for grouping customers, as well as a basis to determine proper allocation of marketing expenditures (Lemon and Mark 2006; Labai, Narayandas and Humby 2002; Bolton, Lemon and Verhoef 2004). Furthermore, CLV based segmentation can also aid marketers in determining what segments they should (or should not) try to create a more profitable relationship with, as well as how to achieve that relationship (Johnson and Selnes 2004; Zeithaml, Rust and Lemon 2001; Libai, Narayandas and Humby 2002; Lemon and Mark 2006; Verhoef and Donkers 2001).

Customer Response Profitability Modeling

According to many researchers, evidence suggests that firms are increasingly relying on long-term customer profitability models to guide marketing strategy decisions (e.g., Libai, Narayandas and Humby 2002; Helf 1998; Peppers and Rogers 1997). Companies are also more and more integrating customer lifetime value (CLV) models into their marketing decisions (such as number of advertisements) in order to ensure that their strategies are effective, thus, optimizing their marketing mix across consumer segments (Libai, Narayandas and Humby 2002; Blattberg, Getz and Thomas 2001; Mulhern 1999; Rust, Zeithaml and Lemon 2000; Zeithaml 2000). However, since CLV modeling research is still in its infancy, researchers and practitioners continue to struggle with how to best assess and manage individual, as opposed to company level, brand level, or product level profitability over time, along with which data are the most valuable in terms of making these predictions (Libai, Narayandas and Humby 2002).

According to Keiningham, Aksoy and Bejou (2006), a serious concerted effort to truly understand individual customer profitability is only a decade old. In fact, they believe, "most firms today still would be hard pressed to say that they have a good understanding of their customers' profitability at the individual level" (2006, p. 37).

These implications are profound! Companies continue to waste precious time and resources in order to target unprofitable accounts. Thus, managers are currently faced with two dilemmas, (1) the fact that CLV and customer profitability research is still in its infancy and is in need of refinement, and (2) how much data and time is really necessary before appropriate targeting decisions can be made? While it is essential to advance the customer profitability knowledge base, in order to successfully do this, it is just as valuable to explain to managers the relative importance of each type of data to their decision making processes.

Current Customer Response Modeling Limitations

Since the majority of CLV modeling is barely more than a decade old, there is much room for improvement. As firms continue to improve their abilities to track and calculate customer lifetime value, there are several issues that need to be better understood. While the topic began to receive attention in the direct marketing literature in the 1990s (Dwyer 1997), only recently have a number of sophisticated mathematical models been proposed to advance the precision of CLV calculation (e.g., Ching, Ng, Wong and Altman 2004; Gupta, Lehmann and Stuart 2004; Jain and Singh 2002; Kumar, Ramani and Bohling 2004; Schmittlein and Peterson 1994; Shih and Liu 2003). CLV models have also advanced very recently in that they have evolved from aggregate measures to individual level calculations (Keiningham, Aksoy and Bejou 2006; Kumar, Ramani and Bohling 2004; Libai, Narayandas and Humby 2002; Venatesan and Kumar 2004; Hogan, Lemon and Rust 2002).

According to Hogan, Lemon, and Rust (2002), a needed change is to develop a truly dynamic individual customer profitability model. Although current CLV models claim to assess the value of a customer, most of the models in the literature still assess the "average" value of a customer and not individual level data (e.g., Blattberg and Deighton 1996; Blattberg, Getz and Thomas 2001). According to Hogan, Lemon and Rust (2002), average valuation models are unable to support the marketing decision making that increasingly occurs at the individual level. Likewise, Libai et al. (2002) indicate that current CLV models also suffer because they do not incorporate any variables other than various profitability statistics. Libai et al. (2002) suggest that researchers should incorporate demographic and psychographic variables into CLV models to further understand customer profitability. They use the example of family lifecycle and explain that it has been demonstrated that the changing needs of the family at various lifecycle stages affect its potential profitability (Javalgi and Dion 1999). Additionally, firms seeking to implement CLV models into their decision making have other challenges as well. According to Lemon and Mark (2006) there are multiple issues with CLV modeling ranging from data and analysis, strategy development, program implementation, and evaluation. According to the authors, while our mathematical computing power has expanded, researchers still need to understand some of the more practical issues dealing with CLV models. For example, what do the models tell managers to do in terms of resources allocation and how should the CLV modeling results affect strategies? More importantly, what type of data is the best to use in order to develop sound analytic tools?

Therefore, the purpose of this study is to evaluate comparative models in order to understand which type of data adds more value to the CLV models. Actual consumer transaction and sales data is termed "hard" data, whereas, individual consumer characteristics such as demographics and psychographics is termed "soft" data.

RESEARCH DESIGN AND METHODOLOGY

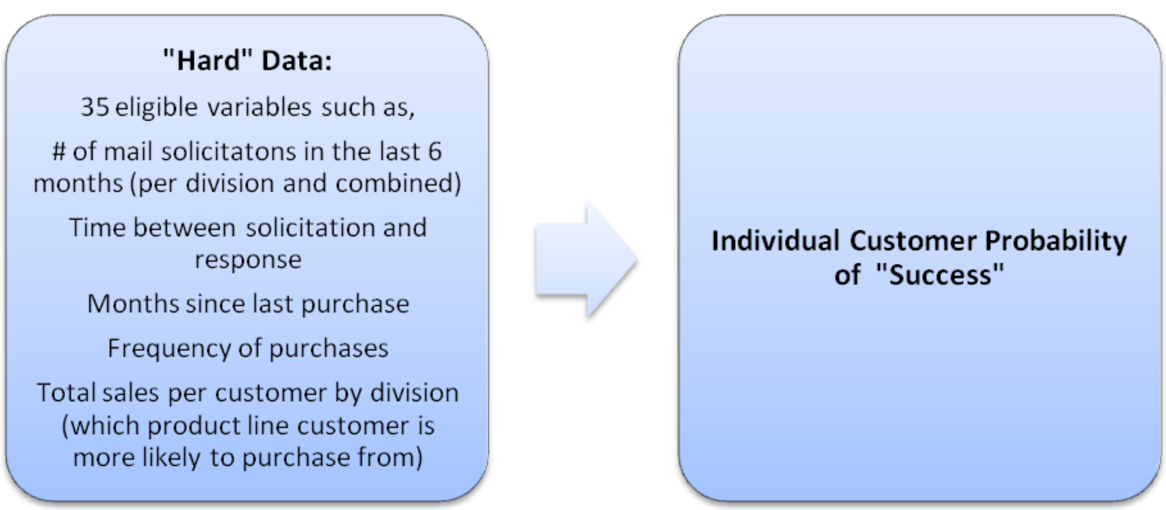
In light of the great importance of both traditional demographic and psychographic segmentation, as well as individual customer profitability response models, it is important that the two fields be integrated to produce even more powerful buyer behavior models. However, the successful combination of the two segmentation approaches is extremely difficult for numerous reasons, including: customer switching/fluctuation/changing (Rust, Lemon and Zeithaml 2004; Reinartz and Kumar 2000; Libai, Narayandas and Humby 2002), different drivers of customer behavior (Lemon and Mark 2006), different behavior patterns/paths in the decision process (Lemon and Mark 2006), the evolution of the relationship between customer and the firm (Johnson and Selnes 2004; Libai, Narayandas and Humby 2002), instability of segment membership (Lemon and Mark 2006), the limitation of mathematical models used to interpret data (Kumar, Venkatesan and Reinartz 2006), firms inability to calculate CLV and inability to forecast customer changes (Lemon and Mark 2006), the minimal correlation between traditional segmentation methods and profitability (Lemon and Mark 2006), the difference in cost of serving customers (Lemon and Mark 2006), the unstable relationship between customer thresholds and customer characteristics (Mittal and Kamakura 2001), and the variation of the dynamics that create value in different industries (Zeithaml et al. 2006).

The integration of analyzing both "hard" and "soft" types of data within the same model is desirable because it will look at between segment heterogeneity and within segment homogeneity. An integrated approach also will allow marketers to examine more dynamic, individual models. Marketers are concerned about getting the most money (or sales) they can for the least amount of expenditures (advertising dollars, product developments costs, etc.) required. With a two level data segmentation approach, marketers can both analyze and target homogeneous segments and find the most preferred segments that maximize profits by looking at within segment profitability and response. The goal though is not to replace traditional segmentation strategies based analyzing only one level of data, rather the plan is to integrate the traditional segmentation approach (i.e., based on soft data) with the customer profitability modeling approach (i.e., hard data). In many academic studies, passing comments have been made about this duel approach (Kumar, Venkatesan and Reinartz 2006; Lemon and Mark 2006; Zeithaml et al. 2006). The closest model found was performed by Mittal and Kamakura (2001) who developed a model that captured the relationship among "rated satisfaction, true/latent satisfaction, repurchase behavior, and consumer characteristics" (Mittal and Kamakura 2001, p.136). Although they provided a useful model to examine the relationship among these factors, they did not look at these factors in relationship to profitability.

What is of even further importance to marketers is understanding which of the two types of data provides the most valuable predictions. Given that marketing budgets are so tight and managers must be accountable for every dollar they spend, it is imperative to not only have a fully integrated model to predict consumer response, but also to have knowledge of the statistical value that each type of data adds to a model. In other words, why pay for demographic and psychographic data, if customer transaction data truly provides the most powerful statistical predictions? Conversely, if consumer transaction data does not do much for prediction, why spend hours combing through it? It all comes down to the age old question of what is best for limited time and limited resources.

For the purposes of this study, the CLV model will enhance existing literature by utilizing individual data for both customer response and for the "soft" demographic and psychographic variables. In addition, the model uses rich customer purchase data rather than dichotomous transaction codes for the independent variables. This is important for managers and researchers alike, who need to understand consumer behavior segments and

FIGURE 1:
Alternative Models for Predicting Customer Response



Model A: "hard" predictors
Model B: "soft" predictors

make important resource allocation decisions within the firm. Also, in addition to refining existing customer response models, this study also builds comparative models in an effort to save executives time and money. As a result, this study postulates the following:

P_1: The added value of the demographic and psychographic data to the customer response model will be insignificant.

These alternative models are depicted in Figure 1.

A secondary outcome will to develop and test a model that investigates how differences in customer characteristics (demographics and psychographics) affect (1) purchase, (2) repurchase, and (3) the likelihood to respond to a direct mailing. Previous segmentation studies show that customers with different characteristics have systematically different thresholds to advertisements and response biases. The extent to which one's purchase behavior, repurchase behavior, and response to ads should vary systematically. Thus, a second postulate is:

P_2: Future profitability can be predicted using (a) past profitability, (b) timing of customer response to direct mailers, and (c) demographic and psychographic variables.

Although there are several challenges with integrating traditional segmentation and customer response data, this research builds upon previous research and offers marketers a more efficient and effective way to tailor their marketing campaigns and allocate resources. In order to do so, a database has been secured of over 175,000 customer purchases over a two year time span. All purchases were made via direct marketing methods. For all consumer responses modeled, the customers were targeted either through direct mail solicitations, Internet web sites, or television infomercials. Purchases were made through the company and sent to the customer through the mail. The purchased products in the sample represent many product categories in consumer durables and non-durables. Products cross many product categories including jewelry, heaters, collectable coins, and diet products. Next, each individual purchase was matched with advertisements received, time from advertisement to purchase, and a large amount of demographic and psychographic data acquired from a major national vendor of data.

To create the dependent variable for the model, a contribution variable was created as a function of backend, lifetime customer sales. To be able to demonstrate P1 we created two categories: customers with contribution, and all other customers. We then modeled the success event. Further models can be developed to tailor our interest on particular customer groups. For example, within the success category we could model the high spenders, or in the failure category we could model the negative group which are customers solicited without any response.

As noted in Figure 1, Model A was created using inhouse, quantitative, independent variables, or so-called "hard" data. Model B utilized the matched demographic and psychographic data acquired. Model C used both the "hard" data and the "soft" data all as predictors of customer contribution. Model A began with 35 eligible independent variables, Model B with 39 eligible independent variables, and Model C with 74 eligible independent variables. (Given the multitude of variables, individual equations are not included in the paper.) Please refer to Tables 1, 2, and 3 for a full list of variables.

ANALYSIS

Strong correlations between variables will result in the existence of multicollinearity and this could cause parameter estimation problems. To analyze multicollinearity, Variance Inflation Factor Analyses were conducted. In the case of logistic regression, values greater than 2.5 could be problematic (Allison 1999). Thus, some variables causing multicollinearity were dropped. The remaining eligible variables had most VIFs ranging around 1.0 or 2.0.

For all models, data were analyzed via a backward elimination logistic regression. This method initially evaluates all the variables, removes the first variable which contributes the least to the R^2, then the method moves to the rest of the variables repeating the first step. To stay in the model, any eligible variable had to meet a $p < 0.05$ significance level equivalent to a 95 percent confidence limit. For the final Model A, 15 variables remained in the model, for the final Model B, 6 variables remained in the model and for the final Model C, 20 variables remained in the model.

In each case, the residual Chi-squared test showed no evidence of model lack of fit with Pr>ChiSq values greater than 0.05. Tables 1, 2, and 3 detail the regression results for Models A, B, and C respectively.

Of importance here is the finding that all the traditionally accepted demographic variables, like age, race, gender, etc., and all of the psychographic variables fell out of this model. The remaining variables all have some association with the person's buying behavior. Chi-squared values show to what extent each variable contributes to the probability being

TABLE 1:
Logistic Regression Results: Model A

		Model A: Hard Data
Independent Variables-	*Predictive Power Rank*	*Estimate*
Lifetime company sales (ltpr)-	3	0.000222**
Months since last purchase from any division (recr)-	1	-0.0610**
Total sales in last 12 months in "c" division (csa12)-	9	-0.00016**
Total sales in last 12 months in "g" division (gsa12)-	14	0.00060**
Total sales in last 12 months in "o" division (osa12)-	11	0.00199**
Months since last purchase from "g" division (grec)-	15	0.00337*
Total sales in last 24 months is "o" division (osa24)-	12	-0.00108**
Number of mail solicitations in last 6 months from "c" division (coff6)-	4	0.0138**
Number of purchases in last 3 months in "g" division (gfreq3)-	8	0.01786**
Number of purchases in last 3 months in "c" division (cfreq3)-	2	0.1752**
Number of purchases in last 3 months in "o" division (ofreq3)-	13	0.5199**
Number of purchases in last 6 months in "o" division (ofreq6)-	5	-0.6527**
Number of purchases in last 24 months in "g" division (gfreq24)-	10	-0.0312**
Number of purchases in last 24 months in "s" division (sfreq24)-	6	0.0742**
Number of purchases in last 24 months in "c" division (cfreq24)-	7	0.0184**
* p < .05		
** p < .01		

TABLE 2:
Logistic Regression Results: Model B

		Model B: Soft- Data
Independent Variables-	*Predictive Power Rank*	*Estimate*
Household Income	*4*	-0.00580**
Household with rental/interest/dividend income-	*1*	0.00512**
Income by ethnicity-	*5*	-0.00017
Those living in homes built prior to 1980-	*3*	-0.00216**
Median year house built-	*6*	0.000332*
Median real estate tax-	*2*	-0.00005**
* p < .05		
** p < .01		

modeled, allowing us to rank them by their predictive power.

The Wald Chi-square estimates give the importance of the contribution of each variable in the model. The higher the value, the more "important" it is. To better understand the magnitude of the differences in those estimates we calculated procentual relative weight values for each variable in Model C ("hard" and "soft" data). The "hard" data accounted for 97.787 percent of the total importance with the "soft" data contributing 2.233 percent of the model's predictive power.

To evaluate how well each model predicts group membership in the dependent variable, we scored the datasets and calculated each decile's predicted probabilities for Models A and C (see Graph 1). The added value provided by Model C when compared to Model A, evaluated from the percent of change point of view, was in the range of only 0.10 to 0.35 percent showing insignificant "soft" data contribution.

These findings strongly support the propositions, which state that "hard" data are much better predictors of future purchases than demographics and psychographics. The question that remains to be answered is whether the small added predictive power of the "soft "data justifies the additional costs associated with acquiring the third-party data and incorporating the information into a segmentation strategy. In other words, would different mail/do not mail decisions be made by incorporating these variables that could be directly attributable to increased profits?

DISCUSSION

The results from our analysis demonstrate several interesting and important findings for both academics and managers; however, the results come with their necessary caveats. All the data used in this study came from a single company. While the purchased products in the sample represent numerous product categories, therefore generalization is of course a concern. Targeting different groups of customers within categories might yield different results.

TABLE 3:
Logistic Regression Results: Model C

		Model C: Hard and Soft Data
Independent Variables-	*Predictive Power Rank*	*Estimate*
Lifetime company sales (ltpr)-	*3*	0.000223**
Months since last purchase from any division (recr)-	*1*	-0.0610**
Total sales in last 12 months in "c" division (csa12)-	*12*	-0.00015**
Total sales in last 12 months in "g" division (gsa12)-	*18*	0.00639**
Total sales in last 12 months in "o" division (osa12)-	*14*	0.00197**
Months since last purchase from "g" division (grec)-	*19*	0.00337*
Total sales in last 24 months is "o" division (osa24)-	*15*	-0.00106**
Number of mail solicitations in last 6 months from "c" division (coff6)-	*4*	0.0137**
Number of purchases in last 3 months in "g" division (gfreq3)-	*10*	0.1691**
Number of purchases in last 3 months in "c" division (cfreq3)-	*2*	0.1747**
Number of purchases in last 3 months in "o" division (ofreq3)-	*17*	0.5171**
Number of purchases in last 6 months in "o" division (ofreq6)-	*5*	-0.6462**
Number of purchases in last 24 months in "g" division (gfreq24)-	*11*	-0.0301**
Number of purchases in last 24 months in "s" division (sfreq24)-	*6*	0.0712**
Number of purchases in last 24 months in "c" division (cfreq24)-	*7*	0.0179**
Household Income-	*13*	-0.00521**
Household with rental/interest/dividend income-	*9*	0.00436**
Income by ethnicity-	*20*	-0.00013
Those living in homes built prior to 1980-	*16*	-0.00118**
Median real estate tax-	*8*	-0.00005**

* p < .05	
** p < .01	

Likewise, caution should be made to generalize the results into any service industry given that all responses were strict product responses.

Also, only main-effects were examined in the model. No additional variables were created such as, ratio, longitudinal, or interaction variables. This was done because the intent was only to provide conceptual protocol and model evaluations to compare and contrast the models. Obviously, the addition of different variables could change the model dramatically. However, this was not the goal of the study.

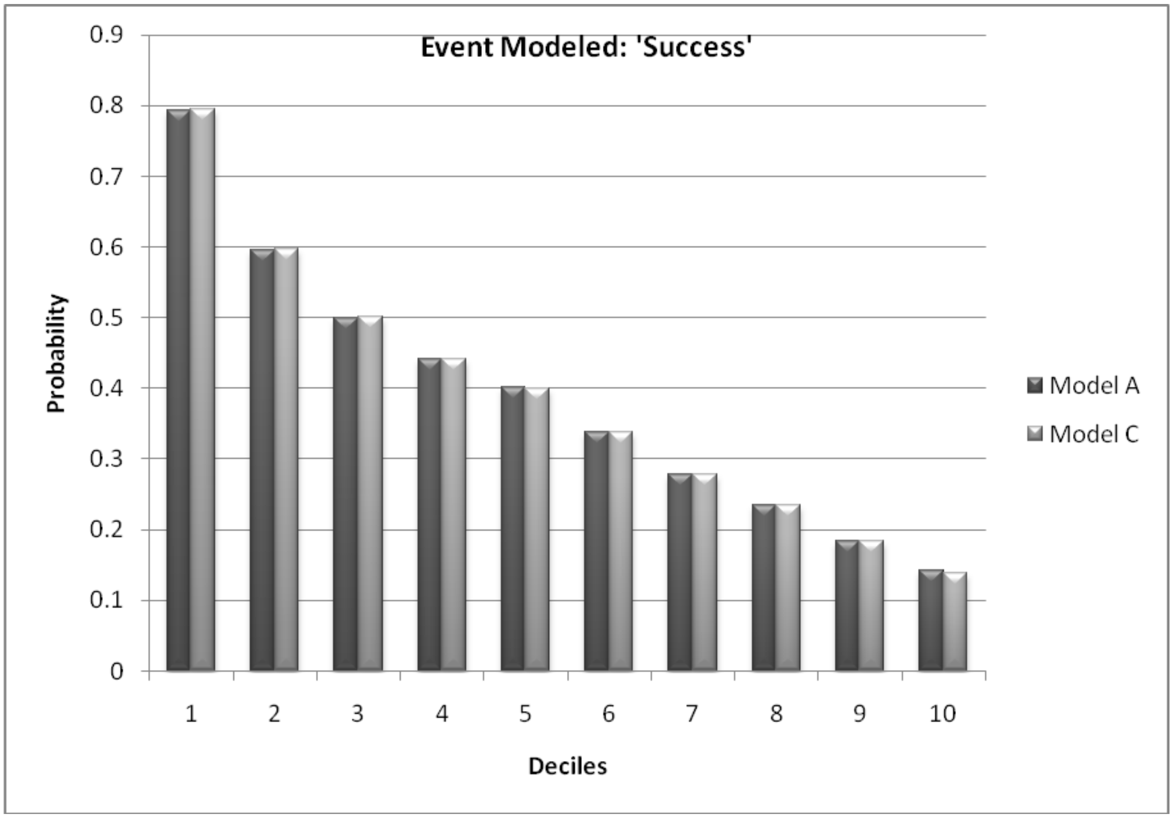
The goal was to examine comparative models to assess the relative importance of adding each

TABLE 4:
Logistic Regression Results: Model C

		Model C: Hard and Soft Data
Independent Variables-	*Predictive Power Rank*	*Wald* X^2 *Relative Weight* *%*
Lifetime company sales (ltpr)-	*3*	13.179
Months since last purchase from any division (recr)-	*1*	54.580
Total sales in last 12 months in “c” division (csa12)-	*12*	.462
Total sales in last 12 months in “g” division (gsa12)-	*18*	.169
Total sales in last 12 months in “o” division (osa12)-	*14*	.308
Months since last purchase from “g” division (grec)-	*19*	.062
Total sales in last 24 months is “o” division (osa24)-	*15*	.308
Number of mail solicitations in last 6 months from “c” division (coff6)-	*4*	9.207
Number of purchases in last 3 months in “g” division (gfreq3)-	*10*	.570
Number of purchases in last 3 months in “c” division (cfreq3)-	*2*	14.257
Number of purchases in last 3 months in “o” division (ofreq3)-	*17*	.200
Number of purchases in last 6 months in “o” division (ofreq6)-	*5*	1.632
Number of purchases in last 24 months in “g” division (gfreq24)-	*11*	.477
Number of purchases in last 24 months in “s” division (sfreq24)-	*6*	1.278
Number of purchases in last 24 months in “c” division (cfreq24)-	*7*	1.078
“Hard” data TOTAL		**97.767**
Household Income-	*13*	.385
Household with rental/interest/dividend income-	*9*	.739
Income by ethnicity-	*20*	.062
Those living in homes built prior to 1980-	*16*	.277
Median real estate tax-	*8*	.770
		2.233

type of data. The contribution of this analysis was to build predictive, more cost effective, customer response profiles for both practitioners and academics with limited resources struggling to better understand how to predict consumer behavior. This goal was achieved in that the proposition in this study was supported. Proposition One stated that the added value of the demographic and psychographic data to the customer response model will be insignificant.

GRAPH 1:
Deciles' predicted probabilities for Models A and C



While many of the variables incorporated into the model predicted consumer response, some definitely predicted better than others. Most notably, the "hard" data represented much by predictors than the "soft" data. This result presents researchers with an interesting thought: accurate customer response models and CLV models can be built using just customer level purchasing transactions history without incorporating demographic and psychographic data. This eliminates additional costs associated with purchasing and processing the supplemental customer data from a third party.

Now, what about a new business venture absent a database with a significant number of customers with a history of customer-level transactions? Researchers can sometimes gain insights into customers from "soft" data. Nonetheless, being able to describe a customer using demographic and psychographic profiles does not imply that better business decisions will be made with this picture in mind, or that more reliable predictions of buying behavior will be made possible by incorporating it. Responsible market testing that generates quantitative (versus qualitative) data can demonstrate the added value of incorporating this information, and should always be performed before incurring the additional costs.

This study also answered the calls by numerous researchers to integrate traditional segmentation models and quantitative consumer response models (e.g., Lemon and Mark 2006; Libai et al. 2002; Hogan, Lemon and Rust 2002; Kumar, Venkatesan and Reinartz 2006). Model C in this study does so. It provides an analytical look at how exactly both hard and soft data can be integrated into a single model to predict consumer response. As noted, this study may have given researchers a different answer than what the research called for. While this study did integrate the two approaches, the findings appear to indicate that this may not be essential as the demographic and psychographic variables appear to have added little in terms of the prediction to the model.

CONCLUSIONS

Business firms and, especially, marketers continuously search for the most predictive models by which to initially segment the customers and, more importantly, maintain profitable relationships with them. The literature in this area strongly suggests the need for additional evidenced based research and model development in this area. With the right amount and type of information, executives can more accurately determine which customers to target with promotions, how often to target them, and even work to understand firm value in terms of combined customer lifetime valuation models.

This study has attempted to present a series of models using actual individual purchase data that has been appended with both customer demographic and psychographic data. The uniqueness of this fully integrated customer database has provided the researchers with the opportunity to test the models in the ideal, and seldom available, real world environment. The findings suggest the strength and robustness of actual consumer purchasing data for use in customer response models over the use of demographic and psychographic data.

However, the research presented here should be viewed as only one attempt in a growing body of research and knowledge building in the area of customer response models and segmentation. At minimum, the findings presented here, should evoke additional research in this area with fully integrated actual customer databases. More importantly, it is hoped that the model building effort here will further add to the discussion and debate over what are the most efficient and accurate ways to develop a customer segmentation strategy.

REFERENCES

Allison, P. (1999), *Logistic Regression Using SAS: Theory and Application*, SAS Press.

Blattberg, R. and J. Deighton (1996), “Manage Marketing by the Customer Equity Test,” *Harvard Business Review* 74(4), 136-144.

Blattberg, R.C., G. Getz and J.S. Thomas (2001), *Customer Equity: Building and Managing Relationships as Valuable Assets,* Harvard Business Press.

Bolton, R., K. Lemon and P. Veroef (2004), “The Theoretical Underpinnings of Customer Asset Management: A Framework and Propositions for Future Research,” *Journal of the Academy of Marketing Science,* 28 (3), 271-292.

Council of American Survey Research Organizations (2009). Available at http://www.casro.org/media/Media%20Facts--US%20(and%20Global)%20Survey%20Research%20Industry.pdf.

Ching, W.K., M.K Ng, K.K. Wong and E. Altman (2004), “ Customer Lifetime Value: Stochastic Optimization Approach,” *The Journal of the Operational Research Society*, 55(8), 860-868.

Dwyer, F.R. (1997), “Customer Lifetime Valuation to Support Marketing Decision Making,” *Journal of Direct Marketing,* 3(4), 8-15.

Edmunds, A. and A. Morris (2000), “The Problem of Information Overload in Business Organisations: A Review of the literature,” *International Journal of Information Management,* 20(1), 17-28.

Gupta, S., D.R. Lehmann and J.A. Stuart (2004), “Valuing Customers,” *Journal of Marketing Research,* 41(1), 7-18.

Helf, I. (1998), “How Segmentation Works,” *Bank Marketing*, 30, 24-30.

Hogan, J.E., K.N. Lemon and R. Rust (2002), “Customer Equity Management: Charting New Directions for the Future of Marketing,” *Journal of Service Research*, 5(1), 4-12.

Hughes, A. (1996), “Boosting Response with RFM,” *American Demographics*, 4-10.

Jain, D. and S.S. Singh (2002), "Customer Lifetime Value Research in Marketing: A Review and Future Directions," *Journal of Interactive Marketing*, 16(2), 34-46.

Javalgi, R.G. and P. Dion (1999), "A Life Cycle Segmentation Approach to Marketing Financial Products and Services," *The Service Industries Journal*, 19(3), 74-96.

Johnson, M.D. and F. Selnes (2004), "Customer Portfolio Management: Toward a Dynamic Theory of Exchange Relationships," *Journal of Marketing,* 68 (April), 1-17.

Keiningham, T.L., L. Aksoy and D. Bejou (2006), "Approaches to the Measurement and Management of Customer Value," *Journal of Relationship Marketing*, 5(2) 37-51.

Kumar V. (2006), "CLV: The Databased Approach," *Journal of Relationship Marketing, 5(2),* 7-35.

Kumar, V., G. Ramani and T. Bohling (2004), "Customer Lifetime Value Approaches and Best Practices Applications," *Journal of Interactive Marketing*, 3, 60-72.

Kumar, V., R. Venkatesan and W. Reinartz (2006), "Knowing What to Sell, When, and to Whom," *Harvard Business Review* (March), 131-137.

Labai, B., D. Narayandas and C. Humby (2002), "Toward an Individual Customer Portfolio Model: A Segment-based Approach," *Journal of Service Research,* 5 (1), 69-76.

Lemon, K. and T. Mark (2006), "Customer Lifetime Value as the Basis of Customer Segmentation: Issues and Challenges," *Journal of Relationship Marketing,* 5 (2/3), 55-69.

Levin, N. and J. Zahavi (1998), "Continuous Predictive Modeling: A Comparative Analysis," *Journal of Interactive Marketing*, 12(2), 5-22.

Marketresearchcareers.com (2009), Available at http://www.marketresearchcareers.com/marketresearchprosurvey2009.aspx.

Mittal, V. and W.A. Kamakura (2001), "Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics," *Journal of Marketing Research,* 38 (February), 131-142.

Mulhern, F.J. (1999), "Customer Profitability Analysis: Measurement, Concentration, and Research Directions," *Journal of Interactive Marketing*, 13(1), 25-40.

Niraj, R., M. Gupta and C. Narasimhan (2001), "Customer Profitability in a Supply Chain," *Journal of Marketing*, 65(July), 1-16.

Peltier, J.W. and J.A. Scribrowsky, J.A. (1997), "The Use of Need-based Segmentation for Developing Segment-specific Marketing Strategies," *Journal of Direct Marketing,* 11 (4), 53-62.

Peppers, D. and P. Rogers (1997), *Enterprise One to One.* New York: Doubleday.

Reinartz, W. and V. Kumar (2000), "On the Profitability of Long-life Customers in a Non-contractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing,* 64 (October), 17-35.

Reinartz, W. and V. Kumar (2002), "The Mismanagement of Customer Loyalty," *Harvard Business Review,* (July), 86-94.

Rust, R.T., K. N. Lemon and V.A. Zeithaml (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing,* 68 (January), 109-127.

Rust, R.T., V.A. Zeithaml and K.N. Lemon (2004), "Customer-centered Brand Management," *Harvard Business Review,* (September), 1-10.

Rust, R.T., V.A. Zeithaml and K.N. Lemon (2000), *Driving Customer Equity: How Customer Lifetime Value is Reshaping Corporate Strategy,* New York: Free Press.

Schmittlein, D.C. and R.A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13(1), 41.

Shih, Y.Y. and C.Y. Liu (2003), "A Method for Customer Lifetime Value Ranking Combining the Analytical Hierarchy Process and Clustering Analysis," *Database Marketing and Customer Strategy Management*, 11(2), 159-172.

Tedlow, R.S. (1990), . *New and Improved: The Story of Mass Marketing in America,* Basic Book Publishers: New York, NY.

Thomas, A.R., L. D. Lewison, W.J. Hauser and L.M. Foley, L.M. (2007), *Direct Marketing in Action: Cutting Edge Strategies for Finding and Keeping the Best Customers,* Praeger: Westport, Connecticut.

Thomas, J., W. Reinartz and V. Kumar (2004), "Getting the Most Out of All Your Customers," *Harvard Business Review*, (July-August), 117-123.

Venatesan R. and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68(4), 106-125.

Verhoef, P.C. and B. Donkers (2001), "Predicting Customer Potential Value and Application in the Insurance Industry," *Decision Support Systems,* (32), 189-199.

Zeithaml, V.A. (2000), "Service Quality, Profitability and the Economic Worth of Customers: What We Know and What We Need to Learn," *Journal of the Academy of Marketing Science*, 28(1), 67-85.

Zeithaml, V.A., R. T. Rust and K.N. Lemon (2001), "The Customer Pyramid: Creating and Serving Profitable Customers," *California Management Review,* 43 (4), 118-142.

Zeithaml, V.A., R.N. Bolton, J. Deighton, T.L. Keningham, K.N. Lemon and J.A. Peterson (2006), "Forward-looking Focus: Can Firms Have Adaptive Foresight?" *Journal of Service Research,* 9 (2), 168-183.