The Dissertation Committee of Shameek Sinha certifies that
this is the approved version of the following dissertation:

Essays in Direct Marketing: Understanding Response Behavior and Implementation of Targeting Strategies

Committee:

Frenkel TerHofstede, Supervisor

Vijay Mahajan, Co-Supervisor

Jason Duan

Romana Khan

Carlos Carvalho

Essays in Direct Marketing: Understanding Response Behavior and Implementation of Targeting Strategies

by

Shameek Sinha, B.Sc.; M.S.; M.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2011

Acknowledgments

I am grateful to Frenkel Ter Hofstede and Vijay Mahajan for their invaluable comments in shaping this dissertation and enhancing its quality. During the course of my dissertation work, they have been a source of inspiration and guidance, pushing me to achieve goals which sometimes seemed impossible to attain. Without their unconditional support, I would not have been able to accomplish what I have so far.

I also thank my committee members Carlos Carvalho, Jason Duan and Romana Khan for their suggestions, insights and encouragement. Also, I would like to acknowledge the role played by Leigh McAlister, Steve Gilbert and David Sibley in providing me invaluable support during challenging times.

I appreciate the role of all my friends, and specially of Dwip Banerjee, Sandip Ray, Shyam Sunder Gauri Suresh, Jacob Brower, Sandeep Arora, Urmee Khan and Priyanka Bardhan who have helped me stay grounded and strong when things seem to drift away.

Last but not the least, I would like to thank my family – my father and mother who had always inspired me in understanding the importance of hard work, moral values and fighting spirit – they have been my friend, philosopher and guide; also my grandmother and grandfather who could not live to see the day I finished my dissertation, but will always be in my heart and mind.

Essays in Direct Marketing: Understanding Response Behavior and Implementation of Targeting Strategies

Publication No.____________

Shameek Sinha, Ph.D.

The University of Texas at Austin, 2011

Supervisors: Frenkel TerHofstede and Vijay Mahajan

In direct marketing, understanding the response behavior of consumers to marketing initiatives is a pre-requisite for marketers before implementing targeting strategies to reach potential as well as existing consumers in the future. Consumer response can either be in terms of the incidence or timing of purchases, category/ brand choice of purchases made as well as the volume or purchase amounts in each category. Direct marketers seek to explore how past consumer response behavior as well as their targeting actions affects current response patterns. However, considerable heterogeneity is also prevalent in consumer responses and the possible sources of this heterogeneity need to be investigated. With the knowledge of consumer response and the corresponding heterogeneity, direct marketers can devise targeting strategies to attract potential new consumers as well as retain existing consumers.

In the first essay of my dissertation (Chapter 2), I model the response behavior of donors in non-profit charity fund-raising in terms of their timing and volume of donations. I show that past donations (both the incidence and volume) and solicitation for alternative causes by non-profits matter in donor responses and the heterogeneity in donation behavior can be explained in terms of individual and community level donor characteristics. I also provide a heuristic approach to target new donors by using a classification scheme for donors in terms of the frequency and amount of donations and then characterize each donor portfolio with corresponding donor characteristics.

In the second essay (Chapter 3), I propose a more structural approach in the targeting of customers by direct marketers in the context of customized retail couponing. First I model customer purchase in a retail setting where brand choice decisions in a product category depend on pricing, in-store promotions, coupon targeting as well as the face values of those coupons. Then using a utility function specification for the retailer which implements a trade-off between net revenue (revenue – coupon face value) and information gain, I propose a Bayesian decision theoretic approach to determine optimal customized coupon face values. The optimization algorithm is sequential where past as well as future customer responses affect targeted coupon face values and the direct marketer tries to determine the trade-off through natural experimentation.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE: INTRODUCTION

Direct marketing has been an area of marketing which has gained a lot of attention from marketing scholars over the last few years. Even though marketers have traditionally focused on customer response behavior to alternative marketing efforts, like pricing, promotions, advertisements etc. and how these marketing strategies can be used to maximize the value of customers through acquisitions, retentions and building loyalty, most of these initiatives have not been linked to direct communications with the customers. Customers' exposure to marketing has been primarily incidental.

However, with the availability of customer level transaction data and other improvements in tracking customers' behavioral patterns, direct marketing seem to have gained momentum as an extremely attractive marketing strategy. Direct marketing can not only help marketers reach customers at the aggregate level, but also can be tuned to reach segments of customers as well as individual customers. Of these possible alternatives, devising direct marketing strategies for individual customers is naturally the most effective mechanism since the marketing communications can be customized to the needs, wants and preferences of each customer. Even though customized direct marketing might be costly to implement, overall the benefits and the incremental value of such strategies usually outweigh the costs, thus proving to be a very viable means of marketing communications and targeting.

Marketing researchers have looked at various aspects of direct marketing. From building customer response models (namely the work by Schmittlein and Peterson (1994), Basu et al. (1995) and Allenby et al. (1999)) which model customer purchase frequency, interpurchase times and monetary value allowing for heterogeneity in customer behavior, to the work in customer relationship management (Reinartz and Kumar (2003), Venkatesan, Kumar and Boehling (2007), Rust and Verhoef(2005), researchers have addressed substantial issues relevant for the marketing practice. Also, there has been a large body of work which adopts decision theoretic approaches towards direct marketing to determine optimal targeting mechanisms (Gonul and TerHofstede (2006), Bult and Wansbeek (1995), Bitran and Mondschein (1996), Gonul and Shi (1998)). However, the direct marketing literature on optimal targeting has been primarily focused on catalog marketing.

Two sectors which have been usually received less attention are the non-profit sector and the retail sector. Even though most of the marketing initiatives undertaken in the non-profit sector are primarily direct marketing, for example reaching potential and existing donors seeking philanthropic contributions for socially relevant causes and also reaching out to volunteers for philanthropic work, this sector has received virtually no attention. This is specially true for the quantitative marketing literature (notable exceptions being Diepen et al. (2009) who focus was on the revenue generation of competing charities and Feder, Hardie and Shang (2010) who model the discrete process underlying donation giving) which can potentially provide significant insights into this sector by building empirical models that help in the understanding of both the marketer's behavior as well as the customer's (in this case the donors and the volunteers). Undoubtedly understanding the response behavior of donors and volunteers in the non-profit sector and deciding on direct marketing policies to reach those groups can prove to be invaluable in a sector where there has often been heated debates regarding possible wastage of resources and inefficient targeting strategies.

The first essay of my dissertation, which I would discuss in Chapter 2 attempts to address some of the relevant issues for direct marketing in the non-profit sector. First, I build a model of donor response behavior to solicitation of funds for philanthropic causes, simultaneously formulating the timing of donations as well as donation amounts. While the timing model considers a multi-episode interval-censored hazard model, the amount model uses a log-Normal model. I explain the heterogeneity in donation behavior using individual and community level donor characteristics with a hierarchical structure in order create donor segments which matches the behavioral pattern of donations to donor characteristics. This not only helps in the understanding of the factors affecting philanthropic donations (specially the relevance of past contributions and solicitations), but also helps in providing insights into the donors segments. I propose a donor targeting mechanism using these donor profiles which can identify expected donation patterns of new donors and future behavior of current donors without even knowing their true donation behavior, but rather by drawing inferences based on the portfolio matching. It is true that this targeting strategy using extrapolation is relatively ad-hoc since I do not explicitly consider the true behavioral patterns.

In essay two of my dissertation (discussed in Chapter 3), where I focus on direct marketing in the retail sector, I provide a theoretical framework to devise an optimal direct targeting strategy after building a model of customer brand purchase behavior influenced by marketing initiatives. The proposed algorithm in that case models optimal decision making as a multi-period sequential problem where information gain and learning during the targeting are the key building blocks.

Though retail marketing has been a prolific research area for marketers, how direct marketing can be effectively used in this sector has a significant void. Except for the work by Rossi, McCulloch and Allenby (1996), who propose a coupon targeting strategy based on a customer brand choice model and analyze the importance of information sets in drawing managerially relevant conclusions, research work on direct marketing in retail is lacking. With retail customers always wanting some additional value for their money, off late direct marketing have gained some popularity in the retail sector as well. Point of purchase couponing and aggregate coupon mailing via newspaper inserts have been a part of the retail business for some time now. Recently, customized direct marketing strategies which are geared towards creating coupons for individual customers with unique face values have started to get some recognition with retailers collaborating with manufacturers. In this context, in the second essay of my dissertation in Chapter 3, I propose a customer response model of brand choice (using a competing risk interval-censored hazard model) by looking at how duration from previous purchases and retailer's brand level targeting strategies using coupons with customized face values affect current purchase of brands. Targeting using coupons by retailers serves two key purposes. First, the targeting by itself serves as an advertisement for the brand and is likely to influence brand choice behavior. Additionally, the customer can choose to redeem the coupon or not redeem the coupon. If he/ she finds the face value of the coupons attractive, then the brand level coupon can be redeemed, thus not only influencing the purchase probability but also possibly lead to brand switching. Pricing across brands within a category as well as other non-price in-store promotional strategies are also considered as factors affecting brand choice.

Then I provide a theoretical framework to target customers in a retail setting using coupons with customized face values. The framework not only allows for a objective function for the retailer which considers the tradeoff between net incremental revenue from the usage of coupons vis-à-vis information gain, but also models the targeting decision problem as sequential, thus providing the opportunity to the retailer to update beliefs over time regarding the customer's response patterns. Computational challenges related to multiplicity of decision paths, evaluation of analytically intractable expected utility integrals and multi-dimensional optimization are addressed using lower dimensional sufficient statistics, forward simulation and grid-based optimization algorithm respectively.

Finally in Chapter 4, I review the theoretical and managerial implications of Essay 1 and Essay 2 in the direct marketing literature since the non-profit and retail sectors are unique. I also point out the possible limitations of the two essays and the potential for future research.

CHAPTER TWO:

EXPLAINING HETEROGENEITY IN DONATION TIMING AND AMOUNT THROUGH INDIVIDUAL AND COMMUNITY CHARACTERISTICS: DEVELOP TARGETING STRATEGIES BASED ON DONOR PORTFOLIOS

Background

With the increase in globalization and information flow, there is an increased awareness regarding income and social disparities, proliferation of political and religious conflicts as well as natural and ecological disasters. Compassionate individuals and socially responsible organizations are keen to get involved in philanthropic activities for the greater good of mankind which has led to a world-wide surge in non-profit charity fundraising for various charitable causes. Non-profit charitable organizations seek donations from current and potential contributors for these philanthropic causes. According to the latest figures from the Survey by The Nonprofit Research Collaborative (conducted by AFP, Guidestar, Foundation Center, Blackbaud, The Center on Philanthropy and NCCS) and the Annual Report on Philanthropy (Giving USA), in 2007 there were almost 1.5 million tax-exempt organizations in the USA which include public charities (64%), private foundations (8%) and other non-profit charity organizations (28%). Public charities, being the largest proportion of non-profit charity organizations, reported around $1.4 trillion in total revenues. While 22% of the revenues came from contributions, gifts and grants, 67% came from program service revenues and 11% from other sources. The reports also reported that the total amount of charitable contributions reached $285 billion in 2008 of which individual donations amounted to $229 billion. Interestingly, around 91% of non-profit charity organizations use direct mail to solicit contributions from donors (2007 figures) with 67% sending more than one mail to the same donors and 24% of these non-profit charities soliciting more than 4 times.

The most pressing issue for the activities of non-profit charity organizations, however, is the proportion of the organization's expenses spent on the actual fundraising activities. Donors are concerned about the non-profit charity's budget spent on administrative expenses as opposed to actual program costs. They expect efficiency on the part of the non-profit charity organization in the collection as well as the disbursement of funds. To provide some benchmarks regarding non-profit charity's fundraising efficiency, watchdog group BBB Wise Giving Alliance recommends that nonprofits spend at least 65% of its annual expenses on program activity, while the American Institute of Philanthropy sets its minimum standard at 60% of expenses. Unfortunately these guidelines are often flouted or sometimes the non-profit charities fail to achieve those targets due to the bureaucratic constraints on the non-profit charitable organization's day to day operations. From 1999 to 2004, researchers at Urban Institute's Center on Nonprofits and Philanthropy and Center of Philanthropy at Indiana University explored issues of non-profit charity fundraising and administrative costs based on surveys of non-profit charities. This "Nonprofit Overhead Cost Study" finds that based on IRS filings, non-profit charities considerably underreport their expenses on fundraising activities for various philanthropic causes. While the non-profit charity organizations emphasize that spending more money facilitates better donations, they also simultaneously ignore some of the overhead costs like value of time from volunteers, efforts by partner organizations, costs of employee and infrastructure dedicated to fundraising etc. Such practices are dictated primarily by the need to tout their low overhead costs ratios in their mailings to the donors, thereby addressing some of the donor concerns.

The above facts bear testament of the importance of the non-profit charity fundraising efforts since these charities not only account for a significant portion of the economy (non-profit charity fundraising accounted for 2.2% of GDP and around 8% of employment in 2008), they raise substantial contributions for various philanthropic purposes. Furthermore, since these non-profit charities also need to improve their efficiency in order to make the fundraising activities worthwhile, direct marketing strategies in this sector assume greater significance. From the non-profit charity organization's perspective it is crucial to understand the contribution behavior of its donors. This provides the non-profit charitable organization the required information to develop efficient targeting strategies for the donors. In this study, we investigate the contribution behavior of donors in terms of their timing of donations as well as the amount of donations. To further enrich the understanding of donation behavior and facilitate efficient fundraising, we explore the role of donor characteristics on the response of donors to targeting efforts of non-profit charities.

The objectives in this research are threefold. First, we formalize the behavioral response of donors to solicitations by non-profit charitable organizations, taking into account past behavior. Second, we investigate the determinants of the donation response behavior, and more specifically the role played by community-level factors. Third, we propose a framework to classify donors in portfolios, which helps non-profit charity organizations to target existing and potential new donors more efficiently.

In order to achieve the first goal of investigating the donation behavior, we simultaneously setup the donation incidence problem using a proportional hazard model with interval censoring which effectively takes the form of a discrete hazard and donation amount problem using a log-Normal regression model. Our work follows the literature on customer response models in the context of direct marketing by building a model of donor responses to direct solicitation efforts in non-profit charity fundraising.

In this research we specifically build on the model developed by Gonul and Ter Hofstede (2006) (albeit in a different empirical context) which allows us to jointly estimate donation response timing and volume decisions of charitable contributions. While they investigate the timing and volume of purchases made by catalog customers taking into account their duration dependence (duration from solicitation and from last contribution), we believe that, in the case of philanthropic donations, the donor's decisions are also driven to a large extent by the time since his last contribution and the time elapsed from the solicitation. Further, the behavioral response of donors is also influenced by their previous contribution amount. Given these considerations, we formulate the donor's decision regarding when to contribute and how much to contribute contingent on his/ her previous donation amounts as well in addition to the duration effects.

To achieve the second objective, namely, capture the heterogeneity in response behavior of donors to solicitation efforts of non-profit charity organizations, we incorporate a hierarchical structure in the behavioral models of donation response. This is formulated such that the response parameters of the donors (in the donation incidence and amount models) for the time-varying covariates (duration from previous contribution, duration from solicitation and previous donation amounts) are functions of donor characteristics at the individual as well as the community level. Note that our emphasis in this hierarchical approach is on the community characteristics determining donation behavior, even though we do explore certain individual level factors. This is because we believe donation behavior is influenced strongly by community environments and social interactions within those communities. Importantly, this hierarchical specification accounts for the observed heterogeneity in the model, whereas unobserved heterogeneity is captured in terms of the diagonal elements of the variance-covariance matrix of the model parameters, which has often been used in the marketing literature (Rossi, McCulloch and Allenby, 1996; Allenby and Rossi, 1999; Manchanda, Ansari and Gupta, 1999).

Last but not the least, as part of the third objective, our aim is to identify donor segments based on their frequency of contributions and the amount of contributions and subsequently to develop targeting strategies for current and potential donors. To accomplish this we conduct predictive validity tests and evaluate the performance of the donation response models. Based on the actual observed data, we first classify donors in terms of their donation frequency and amount. We can readily observe the individual and community level characteristics of the donors in each of these donor groups or portfolios. Then using the estimated parameters for the donation incidence and amount models, specifically those for the characteristics of the donors, we predict the expected probability (frequency) and amount of contributions for each donor. The non-profit organization the matches the characteristics of new donors with their expected behavioral portfolio to determine the donation patterns through extrapolation. Since our aim is to predict donor behavior not only for current donors, but also for potential new donors, we carry out this exercise without explicitly taking into consideration how past donation behavior might affect future behavior. For non-profit charitable organizations, it is often important to acquire new donors for various charitable causes, rather than just concentrating on current donors. For these new donors, no information on previous donation behavior is available which makes the prediction exercise managerially relevant. If this approach is appropriately implemented and provide the expected insightful results, it would help non-profit charity organizations predict donation behavior of communities and individual donors in those communities, without specific knowledge of any past behavior.

In terms of the results, we show that donation behavior to charitable causes has strong duration dependence. Both the duration from last contribution as well as the duration from solicitation strongly affects the incidence of donations as well as the amount of donations. We also stress the importance of past donation amounts in the behavioral response of donors. As we investigate further into the mechanics of donation behavior, we uncover the role played by donor characteristics in explaining how donors respond to solicitations as well as their past donation behavior. With respect to the donor segmentation scheme we propose to target current and potential donors in the future using simply the donor characteristics and not taking into account their past behavior, we had relatively less success. The inaccuracy in the classification scheme with portfolios can be attributed to the sparseness of the donation dataset as well as the noisiness involved in donor characteristics. Unobserved heterogeneity also plays a big role in these results. However, we stress that the approach is theoretically appealing and with better availability of data we can improve on the prediction results considerably.

The chapter is organized as follows. In the next section, we explain in greater detail the mechanisms involved in donation behavior in terms of its duration dependence as well its determinants to set up the problem in hand. In this context we also point out the relevant literature. In the next section, we describe the methodology behind the formulation of the donation response models. We follow that up with a briefly describe the data used for this study along with some descriptive statistics which helps us provide an overall understanding of philanthropic donations and donor characteristics. The next section reports the estimation results from our donation incidence and amount models and explain our findings regarding the role of individual and community level donor characteristics on the behavioral responses of donors. After going over our findings, we explore the donor segmentation approach based on our estimation to target current donors and identify future donors using the predictive validity tests. Finally, we provide some managerially implications for our findings and concluding remarks.

Problem Formulation in the Context of Philanthropic Donations

Donation Response Models

The context of philanthropic donations, the behavioral process underlying donations, its determinants and specially targeting donors with direct marketing efforts, has received relatively little attention in the marketing literature (with the exception of Diepen et al. (2009) and Fader et al.(2010)). However, customer response models in other direct marketing contexts have been extensively studied. The pioneering work by Schmittlein and Peterson (1994) model the expected response behavior of active customers using data on frequency, timing and dollar value of past transactions. Other notable works in this area include Basu et al. (1995) who propose the 'heterogeneous starting point model' to model response pattern to direct marketing campaign and Allenby et al. (1999) who model customer interpurchase times allowing for both temporal and cross-sectional heterogeneity. The response model for philanthropic donation which we use building on the work by Gonul and Ter Hofstede (2006) is drawn from this research stream. The simultaneous consideration of the discrete donation incidence decision and continuous donation amount decision provides a comprehensive framework to formulate the two important aspects of donation behavior.

The customer relationship management literature which looks at the profitability of customer base, also extensively uses response models driven by direct marketing initiatives. Fader, Hardie and Lee (2005) propose a theoretical beta-geometric/ NBD model for predicting future purchase pattern of customers which can serve as inputs to customer lifetime value calculations. While Reinartz and Kumar (2003) explore the role of customer relationship characteristics on profitable lifetime duration, Reinartz, Thomas and Kumar (2005) investigate the optimal balance of scarce marketing resources for customer acquisition and retention in order to maximize profitability. Venkatesan, Kumar and Bohling (2007) propose a joint model of customer purchase timing and quantity and implement a Bayesian decision theoretic approach for optimal customer selection. Rust and Verhoef (2005) suggest a modeling approach to optimally allocate marketing interventions in intermediate term customer relationship management. Our work provides an alternative to these existing approaches to customer value creation in direct marketing. Based on the behavioral response of donors to solicitation efforts by non-profit charities, we create segments of donors with respect to their frequency and amount of donations using individual and donor level characteristics. These donor profiles, when incorporated into the decision making process, increases the efficiency of targeting existing donors in the future and also helps the non-profit charities in identifying and soliciting contributions from potential donors. Therefore, our approach increases the value of current and future donors simultaneously.

Duration Dependence of Donation Behavior and the Importance of Past Behavior

In the field of non-profit charity marketing, Diepen et al. (2009) look at the dynamic and competitive effects of direct mailings on the revenues of each firm involved in soliciting funds for charitable causes, accounting for endogeneity of mailing decisions as well as unobserved heterogeneity. However, their focus is not on the response behavior of donors driven by solicitation activities undertaken by non-profit charitable organizations during the fundraising process, rather on the performance of the non-profit charity organizations. Fader, Hardie and Shang (2010) address this and they propose a beta-geometric/ beta-Bernoulli model to capture the behavioral processes underlying a customer base analysis for donations made to a non-profit charitable organization. They develop a model to predict future donation patterns in a discrete repeat donation setting by taking into account the recency and frequency of donations.

Fader, Hardie and Shang (2010) explain one aspect of the donation behavior, namely the decision regarding whether to donate towards a particular cause. However, when a donor receives a solicitation from the non-profit charity organization asking for contributions towards a philanthropic cause, he/ she has to decide on the amount of donations as well. Therefore incidence and amount decisions are jointly made and need to be investigated simultaneously to get the complete picture. Accordingly, as a first step in this research, we use a framework which takes into consideration the decisions regarding when to donate and how much to donate as the donor actually decides to make a contribution. These two simultaneous decisions made by the donor are contingent upon his/ her previous donation behavior as well as the solicitation strategies used by the non-profit charitable organization to target that donor. While solicitation creates awareness about a particular cause, it might also serve as reminder message for that cause if the last solicitation was some time back. A recent contribution by a donor for a particular cause might constrain him/ her in terms of available budget for philanthropic activities and so delay making another contribution. Further, in philanthropic donations, previous donation amount is important since donors not only have limited budgets for making donations, but also make their decisions after carefully considering their earlier donation amounts for a particular cause. On the one hand, the donor might be reluctant to contribute towards the same cause repeatedly and look for alternative charitable causes. On the other hand, if the donor has contributed a substantial amount of money towards a cause already, it is unlikely that he/ she will be willing to make a large contribution again. Also, if he/ she contributes frequently or contributes towards many causes, his/ her amount of contribution will be lower.

Drivers of Donation Behavior

After analyzing the behavioral aspects underlying donation behavior in terms of incidence of contributions as well as amount of contributions, non-profit charity organizations need to understand the factors driving such behavior. Often the challenge for the non-profit charity organization is not necessarily confined to understanding the donation behavior of its current donors based on their contribution history, but also to predict how potential donors are likely to behave. While current donors are important to non-profit charity organizations because of their generosity towards various ongoing philanthropic causes, non-profit charities also seek to expand their donor base like any other business enterprise for additional contributions towards existing and newer programs. So, uncovering the factors which influence the behavioral aspects of donations provides an understanding of the current donors and helps identify new donors.

Since donation behavior exhibits considerable heterogeneity, efficient targeting by non-profit charitable organization requires the characterization of this heterogeneity and creating donor profiles. Unobserved heterogeneity can be captured without explicitly taking into account additional factors driving donors' behavior. However, one must note that presence of large degrees of unobserved heterogeneity might make drawing managerially interesting conclusions based on the findings, more difficult. Research on the determinants of donation behavior in non-profit charity fundraising, which help explain the observed heterogeneity in donation response to solicitation, has been done in the area of sociology, applied economics and to a certain extent in the nonprofit sector marketing literature. Sargeant (1999) proposes one of the leading conceptual models of the determinants of donor behavior by drawing from the fields of marketing, economics, clinical and social psychology, anthropology and sociology. He argues that after receiving solicitation messages from non-profit charity organizations, donors combine certain intrinsic and extrinsic factors with their perceptual reactions in order to make the decision to contribute. Schlegelmilch and co-authors (Schlegelmilch et al., 1997-I; Schlegelmilch et al., 1997-II) look at the characteristics of donors affecting charitable behavior, including demographic, psychographic and socio-economic factors. Jones and Posnett (1991) explore the determinants of charitable giving in the UK using generalized Tobit models and report the role played by income levels of donors and other demographic variables. Kitchen and Dalton (1990) look at the determinants of charitable donations by families in Canada and using similar methodology as Jones and Posnett find that tax deductions for charitable donations and regional differences explain a lot of the heterogeneity in donation behavior.

In the nonprofit sector marketing literature, Bennett (2003) investigates the underlying values, inclinations and other donor characteristics behind the motivation to donate to particular types of non-profit charity organization whereas Bussell and Forbes (2002) specifically looks at the reasons behind volunteering for charitable causes.

Researchers have also attempted to capture heterogeneity in donation behavior through segmentation schemes. Schlegelmilch and Tynan (1989) try to distinguish donors from non-donors using various demographic, socio-economic, psychographic and situational characteristics and suggest targeting strategies for the most likely donors by identifying distinct donor profiles related to their characteristics. Harvey (1990) identifies five different segments in fundraising using a benefit segmentation approach in terms of demographic characteristics, beliefs, attitudes, as well as certain city-specific descriptors.

Building on this stream of work, the next step of our research focuses on the understanding of the determinants of donation behavior. Equipped with the knowledge about the donor response to solicitations in terms of incidence of contributions as well as amount of contributions and how past behavior plays a role, we explore what factors drive the behavioral processes underlying the contributions towards a philanthropic cause. We believe that the observed heterogeneity in donation behavior can be explained by two sets of factors – first, there are the donor level factors which influence their individual decision making; second, there are the community level factors. Though observed heterogeneity in donors at the individual level can be important in explaining donation behavior, the unique nature of philanthropic donations makes community level variables equally important, if not significantly more crucial, especially when targeting new potential donors. The rationale being, communities can be easily identified and hence more efficient for targeting purposes. This makes creating donor profiles using community characteristics actionable even in large donor databases, whereas individual donor characteristics are not always actionable.

Donation behavior is often the result of social interactions which creates a sense of responsibility, empathy towards others, evokes value judgments, enhances self esteem and facilitates altruistic dispositions. For this reason, the social surrounding of the donors in terms of the economic well being, family structure, political ideology, educational opportunities, religious beliefs, neighborhood safety as well as demographic composition at the community level becomes extremely important in explaining the contribution behavior. So, unlike purchase behavior of standard product categories (for example, customer responses to catalog marketer’s direct marketing efforts), donation behavior towards philanthropic causes is less influenced by the donor’s own characteristics. In addition, most of the research which explores donor level characteristics to explain donation behavior is confined to surveys conducted among potential and existing donors that are obviously limited in their scope and applicability when the direct marketer involved in non-profit charity fundraising has to deal with large donor databases. The psychographic variables, i.e. covariates related to attitudes and beliefs towards altruistic causes, can usually be obtained if a survey is undertaken. Even though the individual level demographic and socio-economic variables, which are usually already available to the non-profit charity in their donor databases, can be useful, it is important for them to combine this information with community level variables since donation behavior is uniquely driven to a large extent by the social interactions between donors within a community. We recognize that the aggregation of information involved in using the community level variables might leave a lot of the heterogeneity in the donation behavior unexplained, but still the importance of these community level factors cannot be ignored. In fact if we can aggregate individual level donor characteristics to construct community level donor profiles, that would not only avoid the problem of mis-aggregation, but also enrich the response models of donation behavior in terms of their accuracy to capture heterogeneity. Researchers in sociology (Hogan and Kitagawa, 1985; Leclere et al., 1998; Sampson et al., 2002), criminology (Peeples and Loeber, 1994; Bellair, 1997) as well as finance (DeMarzo et al., 2004) have tried to understand the role played by community level variables on various outcomes. In non-profit charity fundraising, when community level factors are not available in donor databases, they can be acquired from other secondary sources. With the aggregated community level information, the problems associated with donor level factors in large donor databases can also be addressed.

Classification of Donors for Better Targeting within Segments

Taking into consideration the important role of community level factors, we construct a framework for donation response (incidence and amount) where the focus is on community level demographic composition, family structure, religious inclination, political ideology, educational environment and public safety as the primary determinants of the behavioral response of existing donors. However, from the perspective of the non-profit charitable organization, identifying new donors are equally as important as targeting current donors. We take our research one step forward by proposing a framework which can help non-profit charities construct donor segments in terms of their frequency and amount of contribution. Then we characterize these identified donor portfolios based on the individual and community level donor characteristics which in turn determine the responsiveness of donors to solicitations and past contributions. This characterization can be used to extrapolate the likely donation behavior by matching donor profiles with the timing and amounts of contribution. This not only helps in the prediction of expected behavior of current donors, but also identifies future donors and forecast their donation patterns. This approach attempts to forecast donation behavior and subsequently identify donor groups even in the absence of actual past donation behavior. With this objective in mind, we will first classify the donor behavior according to contribution amount and frequency which can help non-profit charity organizations figure out the distribution of donors on these two dimensions – how often donors are contributing and how much they are contributing (see Table 1). Then we will characterize each of the cells using donor and community characteristics, thus making each segment of donor easily identifiable for the non-profit charitable organization. This two-step approach effectively creates a donor segmentation scheme where the non-profit charities can target the most profitable donors (in terms of frequency and amount) efficiently by understanding who these donors are in terms of their individual and community characteristics.

From the perspective of the non-profit charity organization, Donor Segment D is the ideal segment with donors not only contributing higher amounts every time they contribute, but also these donors make contributions at regular time intervals. On the other hand, Donor Segment A is the least appealing since the donors in this segment contribute lower amounts and that too less frequently. However, Donor Segment B and Donor Segment C are also important. While Donor Segment B comprise of donors who contribute less frequently and higher amounts every time they contribute, Donor Segment C are donors who make donations more frequently, but make lower amounts of contributions. In non-profit charity fundraising, a large proportion of donors are expected to fall in Donor Segments B and C and this makes the segments important to the non-profit charity organization in devising targeting strategies for current donors and potential new donors.

Overall though, the non-profit charitable organization needs to efficiently identify donors in each of the four cells so that they are in a better position to figure out how the donors are responding to their solicitation efforts. Once the non-profit charitable organization can obtain the characterization of the donors, then their decisions are not only driven by the donation behavior, but also by the variation in donor responses which result from the individual and community level donor differences. Giving a ‘face’ to the donors in each of the segments can help marketers devise marketing strategies during solicitation at a relatively aggregate level rather than focusing on each donor individually. Such aggregation might help the non-profit charity organization avoid the perennial problem encountered in charitable fundraising efforts, namely high over head costs which plague optimal disbursement of accumulated donations. Donors who are already in the non-profit charity organization’s database can then be better targeted when seeking additional donations, while at the same time solicitation strategies for potential donors can be effectively devised as well through extrapolation of the donor segments. In this research we provide this donor classification scheme which simultaneously improves the targeting strategies for current and future donors by creating donor profiles. This approach is similar to portfolio analysis and strategic classification approaches used in work by Wind et al. (1983), Armstrong and Brodie (1994) and Grewal et al. (2008), though all three papers looked at these issues in business to business marketing.

We contribute to the understanding of donation behavior by investigating the role of past behavior in new donations and exploring the role of individual and community level donor characteristics in the responsiveness of donors to non-profit charity’s solicitation efforts. We investigate donation behavior both in terms of their timing and amount of contributions. By capturing the observed heterogeneity in philanthropic contributions and proposing the donor classification scheme, we provide a managerially relevant schematic approach whereby non-profit charitable organizations can efficiently target existing donors for additional contributions and also identify potential donors for additional funds. Summarizing, our research addresses the following important issues in non-profit charity fundraising. –

First, we simultaneously investigate incidence and volume decisions in the context of non-profit charitable donation behavior, which has usually been ignored.

Second, duration from last contribution as well as duration from solicitation in addition to past contribution amounts are used in explaining contribution decisions, thus incorporating past history of donation behavior in making current decisions.

Third, the responsiveness of donors in making the donation incidence and amount decisions is formulated as a hierarchical structure where both individual level and community level covariates play important roles. This provides a solution to the large scale observed heterogeneity present in donation behavior for philanthropic causes while at the same time taking into account unobserved heterogeneity through the variance covariance structure.

Fourth, this approach not only helps the non-profit charity organization segment donors efficiently based on their donation behavior and creates donor profiles effectively with the community level characteristics, but also optimize the lifetime value of its donors and create customer equity, which has been an important research theme in the customer relationship management literature.

Fifth, with the understanding of donor portfolios based on donation incidence and amounts and characterized by individual and donor level characteristics, non-profit charity organizations can develop successful solicitation strategies for existing donors as well as potential donors.

Finally, we show that there are discernable differences across different donation programs in terms of the sensitivity of donors to non-profit charity solicitation efforts and past donation behavior, where the heterogeneity in responsiveness to solicitations for donation can be explained largely through individual and community level donor characteristics.

Methodology

In this section, we provide a formal representation of the donation behavior expressed in terms of the donation incidence and donation amount. We build up our model based on framework proposed by Gonul and Ter Hofstede (2006). As explained earlier, donor's behavioral responses in non-profit charity fundraising are driven by two main factors – the solicitation effort undertaken by the non-profit charitable organization seeking donations and the previous incidences of contribution by the donors and the corresponding contribution amounts.

Donation Incidence

As the first step, we formalize the donation incidence as a discrete-time proportional hazard model. We index donors by $i = 1, 2, \ldots n$ who can make contributions towards a philanthropic cause $j = 1, 2, \ldots J$ in periods $t = 1, 2, \ldots T$. Our primary objective is to model the following probability –

$$P\left(y_{it}, z_{it}\right) = P\left(y_{it}\right) P\left(z_{it} \mid y_{it}\right)$$

where $y_{ijt} = 1$ if donor *i* makes a contribution t program *j* in period *t*

= 0 otherwise.

and z_{ijt} : contribution amount of a donor *i* at time *t*.

Note that the probability distribution defined for z_{ijt} is conditional on the occurrence of the contribution incidence defined by y_{ijt} .

For notational simplicity we will not use the subscript referring to particular philanthropic cause (j) in subsequent discussions. For any donor *i*, the hazard of contributing to a particular program is given by –

(1) $$h(d_{it}^c, X_{it}) = \lim_{\Delta t \to 0} \frac{P(d_{it}^c \le T_i < d_{it}^c + \Delta d_i^c \mid T_i > d_{it}^c)}{\Delta d_i^c} = h_0(d_{it}^c) \exp(\beta_i^{/} X_{it})$$

where T_i is the interval between contributions

This proportional hazard function specification above has two components. The first part $h_0(d^c)$ is a time-varying component which captures the dependence of a donation on previous donations in terms of the durations between contributions (d_{it}^c). The corresponding survivor function a time *t* is represented as -

$$(2) \quad S(d_{it}^c, X_{it}) = \exp\left[-\int_0^{d_{it}^c} h_0(u_{it}^c)\exp(\beta_i^{/} X_{it})du_{it}^c\right] = \exp[-\exp(\beta_i^{/} X_{it})H(d_{it}^c)]$$

$$\text{where} \quad H(d_{it}^c) = \int_0^{d_{it}^c} h_0(u_{it}^c)du_{it}^c$$

Even though the donor can make contributions at any time (so the underlying behavioral process is continuous), we can only observe the durations between consecutive donations. Accordingly the data generating process is interval censored. Due to this interval-censoring of hazard data, the discrete analog of the hazard function specification is –

$$(3) \quad \begin{aligned} h(d_{it}^c, X_{it}) &= \frac{S^{/}(d_{it}^c, X_{it})d(d_{it}^c)}{S(d_{it}^c, X_{it})} \\ &\approx P(T_i = d_{it}^c \mid T_i > d_{it}^c - 1) = \frac{S(d_{it}^c - 1, X_{it}) - S(d_{it}^c, X_{it})}{S(d_{it}^c - 1, X_{it})} = 1 - \frac{S(d_{it}^c, X_{it})}{S(d_{it}^c - 1, X_{it})} \\ &= 1 - \exp\left[-\exp(\beta_i^{/} X_{it})\left(H(d_{it}^c - 1) - H(d_{it}^c)\right)\right] \end{aligned}$$

Note that, the discrete hazard function is the conditional probability of observing a donation at time period *t* given no donations has been made till period *(t-1)*. Equation

(3) can be rewritten as – $\log\left(1-h(d_{it}^{c},X_{it})\right)=\exp(\beta_{i}^{/}X_{it})\left(H(d_{it}^{c}-1)-H(d_{it}^{c})\right)$, which gives us –

$$\text{(4)} \qquad \log\left(-\log(1-h(d_{it}^{c},X_{it}))\right)=\beta_{i}^{/}X_{it}+\log(H(d_{it}^{c})-H(d_{it}^{c}-1))$$

The baseline hazard function is then -

$$\text{(5)} \qquad \log\left[-\log(1-h_{0}(d_{it}^{c}))\right]=\log(H(d_{it}^{c})-H(d_{it}^{c}-1))=\log\left[\int_{t-1}^{d_{it}^{c}}h_{0}(u_{it}^{c})du_{it}^{c}\right]=\gamma_{it}$$

Substituting terms in equation (4), we get: $\log(-\log[1-h(d_{it}^{c},X_{it})])=\beta_{i}^{/}X_{it}+\gamma_{i}$ which leads us to the final form of the hazard function as –

$$\text{(6)} \qquad h(d_{it}^{c},X_{it})=1-\exp[-\exp(\beta_{i}^{/}X_{it}+\gamma_{it})]$$

The likelihood for donor *i* making a contribution at period *t* is subsequently written as - $\left(h(d_{it}^{c},X_{it})\right)^{y_{it}}\left(1-h(d_{it}^{c},X_{it})\right)^{1-y_{it}}$

To explore what factors drive the likelihood of making a donation given by the discrete hazard specification, we need to understand the behavioral aspects of donations. First, the timing of previous donations and the corresponding duration between donations affect the probability of making a donation. When a donor makes a contribution towards a particular cause, he/ she is unlikely to make another contribution to that same cause or even a different cause immediately. So in the initial periods immediately following a contribution, the likelihood of another donation may be lower. Also, with limited budget available to the donors of a non-profit charitable organization and their inclination to diversify their donation portfolio between alternative programs, a donor is likely to delay making another donation. Accordingly we define –

(7) $$\gamma_{it} = \alpha_{i0}^{y} + \alpha_{i1}^{y} d_{it}^{c} + \alpha_{i2}^{y} \log(d_{it}^{c})$$

where d_{it}^{c} : duration from last contribution at time *t*.

Here we need to take into account the right-censoring of the donation incidence data in our discrete hazard function specification. Since the donation data is observed discrete spells, the likelihood of observing a data generating process characterized by multiple spells (as is the case for donations where the donor can make multiple contributions within the observation period) is given by the product of the hazard functions with the survivor function. Suppose for any donor *i*, we observe a total of $\tau = 1,...,T_i$ spells. We can rewrite the discrete hazard function as –

(8) $$h(d_{it}^{c}, X_{it}) = P(T_i = d_{it}^{c} \mid T_i > d_{it}^{c} - 1)$$

The survivor function is then the unconditional probability of not observing a donation before d_{it}^{c} and can be written as –

(9) $$S(d_{it}^{c}, X_{it}) = \prod_{u_{it}^{c}=1}^{d_{it}^{c}} (1 - h(d_{it}^{c}, X_{it}))$$

However, the last spell might be incomplete since even if a solicitation is made by the non-profit charity for a philanthropic cause, the observation period might end even before the donor makes a contribution. This is known as the right-censoring problem in the hazard model literature. Let us assume the indicator variable c_i which takes the value 0 if the last spell of donor *i* is censored and 1 otherwise. So if $c_i = 0$, i.e. if the last spell for donor *i* is censored, the likelihood function is of the form - $\prod_{t=1}^{T_i}\left(1-h(d_{it}^{c}, X_{it})\right)$ and if it is completed ($c_i = 1$), the likelihood function is given by -

$$\frac{h(d_{iT_i}^{c}, X_{it})}{1-h(d_{iT_i}^{c}, X_{it})}\prod_{t=1}^{T_i}\left(1-h(d_{it}^{c}, X_{it})\right).$$

Since there is conditional independence in the donation process given the duration between previous donations, the likelihood of observing all donation durations of customer *i* is the product of likelihood functions over all spells, i.e.

$$\text{(10)} \quad \Pr(Donations) = \prod_i \left[\frac{h(d_{iT_i}^{c}, X_{it})}{1-h(d_{iT_i}^{c}, X_{it})}\prod_{t=1}^{T_i}\left(1-h(d_{it}^{c}, X_{it})\right)\right]^{c_i}\left[\prod_{t=1}^{T_i}\left(1-h(d_{it}^{c}, X_{it})\right)\right]^{1-c_i}$$

In addition to the duration from previous contribution, another factor which influences donation incidence is the timing of solicitations made by the non-profit charity organization and the corresponding duration from last solicitation. These solicitations, primarily through direct mail, are in essence similar to direct marketing campaigns undertaken by catalog marketers. When a non-profit charity asks for contributions towards a new philanthropic cause, it serves like a promotional campaign in creating awareness about that cause, thus increasing the likelihood of a contribution.

Alternatively, if donations are sought for an existing program, its effects are not that obvious. Some of these repeat solicitations might create a reminder effect in the minds of the donors, which leads to additional contributions. Unfortunately, the same reminder can sometimes lead to fatigue and thus donors might refrain from contributing to that program. Non-profit charity organizations need to tread on a delicate path when they decide to send out solicitation mails. Also it is important to remember that donors are unlikely to make a contribution immediately after receiving a solicitation. Usually, they spend some time gathering more information about the non-profit charitable organization, specially its reputation, as well as the specifics of the philanthropic cause in question.

Finally, previous donation amounts might also play a pivotal role in the donation decisions. In many cases, the donor chooses to contribute to alternative causes rather than contributing to the same cause. So if in fact a substantial donation is made to a program, the donor is unlikely to contribute to the same program. Even if the donor does not choose to diversify, a big contribution is usually followed by a relatively longer delay before another contribution is made. We include the effects of previous contribution amount in terms a discounted present value of the most current contribution amount. To take into account the effects of these two time-varying factors, we define –

$$\beta_i^{/} X_{it} = \delta_{i1}^{y} d_{it}^{s} + \delta_{i2}^{y} \log(d_{it}^{s}) + \varphi_i^{y} \exp(-r d_{it}^{c}) \tilde{z}_{it} \tag{11}$$

where d_{it}^{s} : duration from last solicitation

$\tilde{z}_{it}$: contribution amount when the last contribution was made

r: discount factor (0.9 in our case following the literature)

Note that we have incorporated the effects of the duration from last contribution (d_{it}^{c}) and the duration from last solicitation (d_{it}^{s}) in terms of a linear as well as a non-linear component (log term). Since we have no prior information regarding the likelihood of donor's making a contribution and its dependence of previous contribution and solicitation behavior, we choose to keep the functional form flexible and the data tell how they affect the behavioral responses. These two terms together will capture possible monotonicity or non-monotonicity. We also emphasize that since donation behavior is highly heterogeneous and we intend to capture the unobserved heterogeneity in terms of the variance-covariance matrix of the parameters and the observed heterogeneity in responses using donor and community level factors, we keep the parameters at the individual level. We will explain the formulation of the hierarchical structure and the related parameterization when we discuss the heterogeneity specification.

<u>Donation Amount</u>

Since the donation amounts are conditional on the fact that the donor actually makes a contribution, i.e. the donation incidence decision, we would expect that the factors which are relevant in the specification of the donation incidence model are also crucial in formulating the donation amount model. Accordingly, the donation amounts are dependent on the timing of previous donations as well as the timing of solicitation for a philanthropic cause. For example, in case the donor has made a substantial recent contribution, it is likely that he/ she would not immediately make another contribution. However, if some additional funds become available to the donor or the donor prefers to contribute smaller amounts of money each time, then we might actually observe another donation to the same program. So the effect of time elapsed from last contribution might work either way. When a non-profit charity approaches a donor with a solicitation message, it might also lead to additional donations, either because it serves as a reminder, or it might actually create new interest for a particular philanthropic cause. With these considerations in mind we specify the donation amount model as follows –

(12) $$[\, z_{it} \mid y_{it}, \mu_{it}, \sigma_z^2 \,] \sim \begin{cases} \log N(\mu_{it}, \sigma_z^2) & \text{if } y_{it} = 1 \\ I(z_{it} = 0) & \text{otherwise} \end{cases}$$ where

(13) $$\mu_{it} = \alpha_{i0}^{z} + \alpha_{i1}^{z} d_{it}^{c} + \alpha_{i2}^{z} \log(d_{it}^{c}) + \delta_{i1}^{z} d_{it}^{s} + \delta_{i2}^{z} \log(d_{it}^{s}) + \varphi_i^{z} \exp(-r d_{it}^{c}) \tilde{z}_{it}$$

The lognormal distribution for z_{it} (with mean μ_{it} and variance σ_z^2) ensures that the donation amount is always positive. Similar to the donation incidence model we include both a linear and non-linear term to explore the effects of the duration variables, namely the duration from last contribution and the duration from solicitation. Since we do not have prior knowledge about the effects of these variables, we keep the functional form flexible and allow the data to tell us the true nature of the effects, namely linearity or non-linearity. Also note that we model parameters at the individual donor level since we investigate the unobserved and observed heterogeneity in donation amount in terms of the hierarchical specification in the next section.

Heterogeneity Specification

The response parameters for the donation incidence and amount models in equations (7), (11) and (13) are donor specific. As we have explained earlier, the heterogeneity in the behavioral responses of donors can be explained through observed donor characteristics, even though some of it might be unobserved. The response parameters might also be correlated across donors. As we have discussed earlier, the presence of unobserved heterogeneity is sometimes critical in the model conclusions and so we would try to understand the extent of this problem when discussing the estimation results. We formulate the observed heterogeneity in donation behavior in terms of the donor-level and community level factors. Even though individual-specific factors are important, as noted earlier, community-level factors play a bigger role in donation behavior. This is because donation behavior is often the result of social interactions between current and potential donors. Also, from a purely marketing point of view, more aggregated community level factors are more actionable when targeting current donors and identifying potential donors.

For the donation incidence model, we incorporate the observed heterogeneity in donor's behavior using the following specification –

$$
(14) \qquad \begin{pmatrix} \alpha_{i0}^y \\ \alpha_{i1}^y \\ \alpha_{i2}^y \\ \delta_{i1}^y \\ \delta_{i2}^y \\ \varphi_i^y \end{pmatrix} \sim N\left(\begin{pmatrix} \bar{\alpha}_0^y \\ \bar{\alpha}_1^y \\ \bar{\alpha}_2^y \\ \bar{\delta}_1^y \\ \bar{\delta}_2^y \\ \bar{\varphi}^y \end{pmatrix} + \Phi^y w_i, \Sigma^y \right)
$$

where w_i are donor-level and community-level factors influencing donation incidence, Φ^y is the vector of parameters for the covariates and Σ^y is the variance-covariance matrix for the time-varying factors. Note that the diagonal elements of Σ^y are used as estimates of unobserved heterogeneity in donation incidence.

In a similar way, the heterogeneity exhibited in donation amount is modeled as -

$$
(15) \qquad \begin{pmatrix} \alpha_{i0}^z \\ \alpha_{i1}^z \\ \alpha_{i2}^z \\ \delta_{i1}^z \\ \delta_{i2}^z \\ \varphi_i^z \end{pmatrix} \sim N\left(\begin{pmatrix} \bar{\alpha}_0^z \\ \bar{\alpha}_1^z \\ \bar{\alpha}_2^z \\ \bar{\delta}_1^z \\ \bar{\delta}_2^z \\ \bar{\varphi}^z \end{pmatrix} + \Phi^z w_i, \Sigma^z \right)
$$

where Φ^z is the vector of parameters for the covariates influencing donation amount and Σ^z is the corresponding variance-covariance matrix, which again provides an estimate of unobserved heterogeneity in donation amount.

Estimation

The estimation for both the donation incidence model and the donation amount models are done using Markov Chain Monte Carlo (MCMC) methods. Note that the estimation is done using mean-centered data, the time-varying covariates are mean-centered with respect to their donor-level means and the donor characteristics are mean-centered with their means across donors. In Bayesian MCMC methods, posterior distributions are approximated by sampling from the full conditional distributions where full conditionals are obtained from the prior distribution of the parameters and the observed data using Bayes rule. In the estimation of the donation incidence model, we do not have closed form representations of the full-conditional distributions for the donor-specific parameters $\{\alpha_{i0}^y, \alpha_{i1}^y, \alpha_{i2}^y, \delta_{i1}^y, \delta_{i2}^y, \varphi_i^y\}$. So we use the Metropolis-Hastings algorithm with a Normal distribution for the proposal densities involved in the Metropolis-Hastings steps. However, for the parameters of the donation amount model $\{\alpha_{i0}^z, \alpha_{i1}^z, \alpha_{i2}^z, \delta_{i1}^z, \delta_{i2}^z, \varphi_i^z\}$, we can obtain exact full conditional distributions and accordingly we use the Gibbs sampler to approximate the posterior distributions.

The response parameters are donor specific which allows for different response patterns across donors. While unobserved donor characteristics might affect the shape of these curves, the response parameters may also be correlated. So we assume the following distributional specifications for the donor-specific parameters which allows for the unobserved factors –

$$[\alpha_{i0}^y, \alpha_{i1}^y, \alpha_{i2}^y, \delta_{i1}^y, \delta_{i2}^y, \varphi_i^y]' \sim N([(\bar{\alpha}_0^y, \bar{\alpha}_1^y, \bar{\alpha}_2^y, \bar{\delta}_1^y, \bar{\delta}_2^y, \bar{\varphi}^y, \Phi^y)]', \Sigma^y)$$

and $[\alpha_{i0}^{z},\alpha_{i1}^{z},\alpha_{i2}^{z},\delta_{i1}^{z},\delta_{i2}^{z},\varphi_{i}^{z}]^{/} \sim N([(\overline{\alpha}_{0}^{z},\overline{\alpha}_{1}^{z},\overline{\alpha}_{2}^{z},\overline{\delta}_{1}^{z},\overline{\delta}_{2}^{z},\overline{\varphi}^{z},\Phi^{z})]^{/},\Sigma^{z})$

Note that the posterior distributions of the donor level parameters for the donation incidence model are approximated using the Metropolis-Hastings algorithm while those for the donation amount model are approximated using the Gibbs sampler.

The prior distributions for the population level parameters $\{(\overline{\alpha}_{0}^{y},\overline{\alpha}_{1}^{y},\overline{\alpha}_{2}^{y},\overline{\delta}_{1}^{y},\overline{\delta}_{2}^{y},\overline{\varphi}^{y},\Phi^{y},\Sigma^{y});(\overline{\alpha}_{0}^{z},\overline{\alpha}_{1}^{z},\overline{\alpha}_{2}^{z},\overline{\delta}_{1}^{z},\overline{\delta}_{2}^{z},\overline{\varphi}^{z},\Phi^{z},\Sigma^{z})\}$ are taken from standard conjugate hyperdistributions, namely –

$$(\overline{\alpha}_{0}^{y},\overline{\alpha}_{1}^{y},\overline{\alpha}_{2}^{y},\overline{\delta}_{1}^{y},\overline{\delta}_{2}^{y},\overline{\varphi}^{y},\Phi^{y})^{/} \sim N((\alpha_{00}^{y},\alpha_{01}^{y},\alpha_{02}^{y},\delta_{01}^{y},\delta_{02}^{y},\varphi_{0}{}^{y},\Phi_{0}{}^{y})^{/},\Lambda_{0});$$
$$\Sigma^{y} \sim Wishart(\lambda_{0}^{y},\Sigma_{0}^{y})$$

$$(\overline{\alpha}_{0}^{z},\overline{\alpha}_{1}^{z},\overline{\alpha}_{2}^{z},\overline{\delta}_{1}^{z},\overline{\delta}_{2}^{z},\overline{\varphi}^{z},\Phi^{z})^{/} \sim N((\alpha_{00}^{z},\alpha_{01}^{z},\alpha_{02}^{z},\delta_{01}^{z},\delta_{02}^{z},\varphi_{0}{}^{z},\Phi_{0}{}^{z})^{/},\Lambda_{0});$$
$$\Sigma^{z} \sim Wishart(\lambda_{0}^{z},\Sigma_{0}^{z})$$

The full conditional posterior distributions of these population level parameters reduce to Normal, Gamma and Multinomial distributions. For the donation incidence and amount models, we made 40,000 draws from the full conditional distributions to approximate the posterior distribution with 15,000 draws as burn-in to ensure better convergence. Also in order to avoid poor mixing resulting from high sample autocorrelation, we used a thinning parameter of 20, i.e. we retained every 20-th draw.

Empirical Application

Data

Our donation data comes from a non-profit charitable organization that uses direct mail to solicit additional contributions from its existing donors. Even though we had data available for the entire USA, we focus on the data specifically for the state of Texas, mainly because we wanted to keep the analysis manageable since Bayesian MCMC methods are usually very computation intensive specially in the simulation stages while at the same time explain the donation behavior, its underlying factors and also capture sufficient degree of heterogeneity in donor responses in Texas. Data was provided to us by Direct Marketing Education Foundation (Code: 01DMEF). The data covered approximately a 9 year window (October, 1986 to June, 1995). We have information on the timing and amount of contributions as well as the timing of solicitations by the non-profit charity organization in that period. We have two different philanthropic causes for which the non-profit charity organization seeks contributions from donors. However, we do not have information on the actual nature of the causes due to privacy concerns, which somewhat restricts our interpretation of the parameter effects. This data is used to estimate the donation incidence and amount models.

Now in order to formulate the observed heterogeneity in response behavior, we need additional donor-specific and community level variables which explain the differences in behavioral responses of donors. The donation database we use for the response models have only limited donor-specific covariates, more specifically we only have the gender of the donor and number of lifetime solicitations by the non-profit charitable organization for each donors, available. So we append our dataset with two additional datasets available from DMEF. One of them ZIP code level geo-demographic data and other has ZIP code level credit information. Since this datasets are aggregated at ZIP code level, it serves our purpose of focusing on community level donor characteristics in explaining the heterogeneity in donation response. From the geo-demographic dataset, the variables which we believe to be potential factors influencing donation behavior are – 1) Median age of households; 2) Percentage of white in the population; 3) Median years of schooling; 4) Persons per household; 5) Percentage of households which are families; 6) Estimated median income and 7) Median home value. From the credit dataset we use total number of tradelines as a factor which approximates the financial strength and stability of the donor household.

As we have emphasized earlier, in non-profit charity fundraising community factors better explain the social interactions at work and so we use four additional data sources to incorporate the effects of other important community features. These are –

1) The Texas Education Agency's (TEA) Accountability Rating Systems for Schools at the ZIP code level is used to incorporate the effect of quality of education since community education standards might affect both the incidence and amount of donations. These ratings are measured by the TEA for every school as a weighted average of Texas Assessment of Knowledge and Skills (TAKS) scores and dropout rates. We aggregated the school-level ratings to the ZIP code level.

2) Information on the number of churches at the ZIP code level for the different Christian beliefs across Texas is analyzed to understand the effects of religious inclinations on donation behavior. Even though we had the data on the number of churches per capita available for each of the different Christian beliefs, since a large portion of churches in Texas were Protestant, we decided to make the distinction in religious ideologies with respect to Protestants versus Non-Protestants.

3) Percentage of votes won by political parties in a Presidential election at the county level is used as a proxy for the political beliefs in Texas and this might explain some of the heterogeneity across communities. Since the electorate in Texas only comprised of Democratic and Republican party voters (insignificant number of Independents), we decided to use percentage of Republican votes in Texas as proxy measure of political ideology.

4) Finally, county-level crime statistics obtained from FBI Crime in the US (CIUS) datasets is included to investigate whether level of public safety in a community matters for philanthropic donations. Even though the FBI crime statistics database includes detailed information on all types of crimes at the county-level, we decided to use the aggregate of total number of violent and non-violent crimes per capita as a measure of public safety.

Summary Statistics

In Table 2 we provide some descriptive information about the two donation programs in our dataset in terms of their estimation period characteristics. From the table, we see that there are some important differences between the two programs we consider in our empirical application. While donors on average contribute more times to donation Program 1 compared to Program 2 (3.99 times versus 1.83 times), they usually contribute less every time to Program 1 compared to Program 2 ($8.97 against $13.58). It also turns out that the duration between consecutive contributions to Program 1 is considerably lower (21.95 weeks) compared to Program 2 (39.36 weeks). The corresponding frequencies of donations are also significantly different, 0.063 and 0.036 respectively. So, even though contributions to Program 2 are done less frequently, every time a contribution is made, it is more substantial. This might result from the higher number of solicitations on average for Program 1 compared to Program 2 (5.30 against 2.98). Interestingly, once a non-profit charity solicits for donations for a particular program, donors usually take similar lengths of time to respond to those solicitations for either program. These facts might imply that donors show more interest towards Program 1 by contributing often, but at the same time they also feel the need to contribute higher amounts to Program 2, when they actually decide to contribute to that program.

The next logical question to address is – what might explain the differences in the behavioral responses of donors across the two donation programs. In order to get a better understanding of the individual-level and community level factors which potentially explain the observed heterogeneity in behavioral responses of donors, Table 3 reports some descriptive statistics of these factors contrasted across the two donation programs. From the table we see that even though there are discernable differences between the two donation programs with respect to the behavioral response of donors and also the solicitation effort made by the non-profit charity organization, on average the individual and community-level donor characteristics are similar. We do observe some differences mainly with respect to the gender composition, number of lifetime solicitations, household size, median home value and the per capita crime rate. So on average, across the two donation programs, individual and community-level characteristics of donors might not be sufficient to explain the differences in donation behavior. A considerable portion of the heterogeneity in their behaviors might be unobserved, and so we need to consider that while drawing conclusions. However, note that these statistics are aggregated across ZIP codes and counties. Also we do not make distinction between donors who differ in terms of their frequency of donations as well as donation amounts. For that purpose, we refer back to Table 1. In that table we proposed a classification scheme for donors in terms of their donation behavior. Here we replicate that table as Table 4a and 4b (for Program 1and 2 respectively) with additional information on how donors are distributed in each of the four cells of that table and how their donation behavior differs. Then we investigate whether indeed these differences among donors in their contribution behavior can be explained using the individual level and community level characteristics (results reported in Tables 5). In order to classify donors, we used mean frequency of contributions and mean donation amounts in each contribution instance which has been already reported in Table 2. From the results in Tables 4a and 4b, it is evident that donors across the four donor segments differ considerably in their contribution behavior. For both the programs, majority of the donors are 'low amount' donors. These 'low amount' donors donate just a third of the donations made by 'high amount' donors every time they contribute. Also, 'high frequency' donors contribute almost twice as frequently as 'low frequency' donors. Subsequently, the duration between each contribution for these 'low frequency' donors is much higher than the 'high frequency' donors.

With frequency of donations on the horizontal axis and amount of donations on the vertical axis, to understand whether the differences between the number of low and high frequency donors and low and high amount donors are statistically significant, we use Chi-squared tests in the essence of contingency tables. We compare the expected and observed frequencies of donors for each program and find that for Program 1, the Chi-squared test is significant which gives evidence of heterogeneity in donation behavior and the dependence between donation frequency and amount (Program 1 has a value of $\chi^2 =$ 19.46 which is greater than the critical value of 3.84). However for Program 2, we have a value of $\chi^2 = 1.95$ which is less than the critical value, thus leading us to the conclusion that there might not be significant differences among donors in each segment for donation Program 2.

With donors exhibiting significant differences in terms of their frequency and amount of contributions for Program 1, for the rest of the discussion in this section we would specifically focus on Program 1. Since the individual and community-level donor characteristics are also the basis for our predictive validity analysis where we predict donation frequency and amount of donations for each donor segment using these donor characteristics, the discussion on donor classification using the predictive validity exercise will also be concentrated on Program 1 (results for all subsequent analysis for Program 2 are available from the author on request).

First, we try to understand whether this observed heterogeneity can be explained in terms of the individual-level and community-level donor covariates. Once we are able to create donor profiles for each of the donor segments, non-profit charities can not only target current donors more efficiently but also identify as well as solicit donations from potential donors. In Table 5, we report the average values of the individual and community level donor characteristics for each segment of donors for Program 1. Even though there are no distinguishable characteristics between 'low frequency' and 'high frequency' donors, we observe certain differences across 'low amount' and 'high amount' donors. Whereas 'high amount' donors are have higher proportion of male donors, have higher median income and home values as well as higher tradelines and have been solicited more often by the non-profit charitable organization, 'low amount' donors are relatively older, have lower years of schooling, belong to communities with lower school ratings, have a higher share of Republican votes and has higher number of Protestant churches per capita. In order to specifically identify the unique characteristics of these segments, and more specifically to investigate whether these means differ across the groups significantly, we use a two-way ANOVA for each group of donors for every donor characteristic. The two factors were the frequency of contributions and the amount of contributions. We also included an interaction effect in order to identify the potential dependence between the frequency and amount of contribution for each donor segment. Interestingly, the results of the ANOVA provide evidence of the importance of every donor feature we analyzed in explaining the heterogeneity in donation behavior since not only are the main effects of the factors significant, but also the interaction effects are significant. The means of the donor characteristics are in fact different across the four donor groups. This provides empirical justification for using the donor characteristics in our donation incidence and amount models (in a hierarchical structure explaining donor-level parameters). The dependence across the duration between contributions and the amount of contributions also justifies the inclusion of lagged donation amount in our framework.

Estimation Results

The results of the Bayesian estimation for the donation incidence and amount models are reported here. In this section we first discuss how the behavioral response of donors is affected by the time-varying factors, namely the duration variables and past donation amounts (Table 6), and try to provide some intuition for the effects we observe. The parameters discussed are the point estimates obtained by averaging donor level estimates and taking the median values of draws from the posterior distribution. We will follow that up with a discussion of the reliability of the parameters across donors. This is done by looking at the credible intervals which provide upper and lower bounds on the marginal hazard probabilities and marginal donation amounts. These marginal effects are defined in terms of the duration from last contribution as well as the duration from solicitation. In order to visually illustrate the effects of the time-varying duration variables on the hazard of making a donation and the corresponding amount of contributions, we use Figures 1 and 2 where the donation probabilities and amounts are shown along with their reliability bounds for each of the programs. Subsequently, we try to understand the unobserved heterogeneity in donation behavior. As mentioned earlier, we focus on the variance-covariance matrix of the parameters for the duration variables and past donation amount whose diagonal elements give a measure of the unobserved heterogeneity. We will discuss the unobserved heterogeneity in the donation incidence and donation amount models for both programs (Tables 7a and 7b). Then we will explore how the observed heterogeneity in donation behavior can be explained using individual-level and community-level donor characteristics (Table 8). The significance level of the parameter estimates are analyzed in terms of a Bayesian analog of frequentist p-values. According to Gill (2008), given a model, a frequentist p-value is the density under the assumed true null hypothesis starting at the test statistic and continuing up to infinity on the support for the parameter in question, which yields an average over unlikely sample values that have not actually occurred. The Bayesian analog of frequentist p-value is the slice of the density that corresponds to the one-sided restriction defining the null hypothesis calculated over the posterior distribution. Gill (2008) suggests that this Bayesian posterior probability "…is far more useful because it is the value that many people mistake a p-value for: the probability that the null hypothesis is true, given the data and the model". The probabilities are measured as the expectation of observing the parameter across all sweeps in the Bayesian simulation.

Duration Dependence of Donation Behavior

Let us first focus on the duration dependence of donation incidence and donation amount. As we have argued earlier, both the duration from solicitation and from previous contribution are expected to play important roles in the behavioral responses of donors. In case of donation incidence, from Table 6 we see that the duration variables have significant effects on the incidence as well as the amount of donations for both donation programs 1 and 2. In fact there is non-linear effect of solicitation to contribution duration as well as the contribution to contribution duration. To understand the effects of the duration variables more carefully, we plot the marginal effects for both donation incidence and amount models for each of the donation programs under consideration (in Figures 1 and 2). Note that the solid curves denote the median effects of the parameter estimates evaluated over the sweeps across individuals, while the dashed curves represent the 95^{th} and 5^{th} percentile effects, thus providing the 90% credible intervals for the reliability of the parameters.

In Figure 1, we begin our discussion of the effects of the duration variables, namely the duration from last contribution and the duration from solicitation, on donation incidence. While the top panel shows the plots for Program 1, the bottom panel looks at the corresponding figures for Program 2. For Program 1, we see that as the duration from last contribution increases, the likelihood of the donor making another contribution gradually increases. This is because, once a donor makes a contribution, he/ she is initially reluctant to make another contribution because of scarcity of funds available for philanthropic activities. Also, the donor might be interested to look for other donation programs, rather than contributing to the same program repeatedly. As the duration increases however, the donor is more likely to make another contribution. Now if we look at the effects of the duration from solicitation on the probability of making a contribution, we actually see that initially there is a decline in the probability. This is somewhat surprising, but can be the result of the fact that, a donor might be unwilling to make a donation towards a program immediately after receiving a solicitation from the non-profit charity organization even if the solicitation serves as a reminder or creates awareness. He/ she is going to research on the specific nature of the program as well as the profile of the non-profit charity as well. Once the donor is convinced about the worthiness of the donation program, then only a contribution is likely to be made. Accordingly we see that after about 8 weeks, the probability of making a donation increases gradually.

If we compare the effects of the duration variables on donation incidence between Program 1 and Program 2, we notice that even though they look similar overall, there are a couple of interesting differences. The obvious difference is seen in the effect of duration from solicitation on the likelihood of contribution. While for Program 1we have seen a curve which was decreasing initially and then increasing after a few periods, for Program 2 this effect is absent. Even though for the first couple of periods the probability of making a donation has a slower rate of increase, it picks up pace after a around week 4 and then gradually increases in a similar fashion as Program 1. The second difference between Program 1 and Program 2 is seen in the effect of duration from previous contribution. Though the curves looks similar, on closer observation it is revealed that for Program 1, the probability of making a donation is 0.03 in the period following the previous contribution and then it increases to around 0.33 at the end of 26 weeks. On the other hand, for Program 2, the likelihood of a contribution following the last contribution is almost negligible. However, the rate of increases is much faster and reaches around 0.48 at the end of 26 week. This observation means that donors of Program 2 are usually more likely to make a contribution as the duration increases. A similar pattern of results is seen for the duration from solicitation as well, where we see that the rate of increase in the hazard of making a contribution is higher for Program 2 compared to Program 1.

Now we compare the effects of the duration variables on donation amount, which is conditional on the fact that the donor actually makes a contribution, in Figure 2. Again we compare results between Program 1 and Program 2. For Program 1, for the first 2 periods we see that the amount of a second contribution, given the donor actually decides to donate again, is actually lower. This is intuitive since the budgetary constraints faced by a donor are usually more important in the periods immediately following a contribution. This is not the case as the duration increases, and thus we observe a curve which is increasing gradually. For the solicitation to contribution duration, we simply see an increasing effect, which shows that once a non-profit charitable organization seeks donations from potential contributors, they are going to reciprocate by making contributions which increases over time resulting from a feeling of social responsibility on the part of the donor. In case of Program 2, the effect curves on donation amounts are increasing for both duration variables. However compared to Program 1, the rate of increase is much higher in both cases. While it picks up pace after around 3 weeks for the duration from previous contrition, the rate of increase in donation amount due to the duration from solicitation is significantly higher after 6 periods. Comparing Program 1 and Program 2 also leads us to make another interesting observation. The changes in donation amounts driven by the duration variables, is greater for Program 2 compared to Program 1 as can be seen from the large scale differences on the vertical axis for the two programs. Also on aggregate the duration from solicitation leads to a greater increase in donation amount over time. Even though the increase in duration from last contribution leads to an increase in the donation amounts, the duration from solicitation has a greater effect because donors are usually more likely to react favorably to a solicitation effort from the non-profit charity organization. A solicitation creates awareness, feeling of empathy, guilt and social consciousness in donors to get involved in philanthropic causes, which is less likely to happen spontaneously out of self exploration.

<u>Reliability of Parameters Exhibiting Duration Dependence</u>

We now look back at Figures 1 and 2 to explore the reliability of the effects which we discussed above. Reliability of parameter estimates is extremely important, specially in Bayesian estimation methods, because unless we can provide results which are consistent and stable, the conclusions drawn in the managerial decision process might be flawed. Overall we see that for both Program 1 and 2, the effects of duration from last contribution as well as the duration from solicitation are extremely reliable, as illustrated in Figure 1. The 5% and 95% bounds are pretty tight and show the stability across the draws from the posterior distribution of the parameters. We do see some widening of the bands as the durations increase because of inherent heterogeneity in the donation behavior, both unobserved and observed.

The reliability of the effects of the time-varying variables on donation amount is much less reliable, specially for the donation Program 2. It is clear from Figure 2 that the 5% and 95% bounds for the effects of duration from contribution as well as solicitation are very much within acceptable limits for Program 1. For Program 2 though, they are really unstable which makes drawing conclusions regarding the effects on donation amount of the duration variables relatively tricky and the managers should be circumspect in making any concrete decisions based on these results. We would like to point out here that the observations we made regarding the differences in Program 1 and Program 2 with respect to the effects of the duration variables on donation incidence, donation amount as well as the reliability of the effects, can be attributed to the very nature of the programs themselves. Due to privacy restrictions of the donors, we were not aware of the actual philanthropic causes which were the motivations behind the two programs, and so the true explanation of these differences remains unexplored.

Explaining the Heterogeneity in Donation Behavior

Unobserved Heterogeneity:

First we explain the extent of unobserved heterogeneity present in the donation incidence and amount models. In Tables 7a and 7b we report the variance-covariance matrices for the incidence and amount models and focus our attention on the diagonal elements since these give us the degree of unobserved heterogeneity present in the estimated parameters.

From the tables it is clear that both the donation incidence and amount models for Programs 1 and 2 exhibit considerable degree of unobserved heterogeneity. This can be easily seen when we consider a couple of examples. In the donation incidence model for Program 1, the coefficient of the log effect of duration from solicitation is -0.135 with a standard deviation of 1.637 and the coefficient of the log effect of duration from last contribution is 0.415 with a standard deviation of 0.719. Both of these show large amount of donor level unobserved heterogeneity. The discounted value of past donation amount also shows extensive heterogeneity since it has a parameter of -4.861 with a standard deviation of 3.4. Almost similar pattern of donor level heterogeneity is seen in the incidence model for Program 2. When we look at the donation amount model for Program 1, again unobserved heterogeneity seems to be present for both the linear and non-linear effects of the duration variables as well as for the effect of the discounted value of past contributions. For example, duration from last contribution and its log have coefficients 0.033 and -0.409 respectively with the corresponding standard deviations given by 0.015 and 0.216. Results look similar for Program 2. It is important to note that this unobserved heterogeneity at the donor level is conditional on the hierarchical structure we formulated where observed heterogeneity is explained in terms of community level donor characteristics. This means that even after covering for the observed heterogeneity, a substantial amount of the heterogeneity in donor behavior will remain unexplained.

Observed Heterogeneity:

With the hierarchical framework we use for the donation incidence and amount models, we can investigate how the individual and community-level (ZIP Code and county) donor characteristics affect these two decisions, thus explaining the observed heterogeneity in the donor's response behavior. More specifically, we try to understand whether the responsiveness of donors with respect to the past donation and solicitation behavior are determined based on which communities they stay in, which demographic and financial characteristics are relevant, whether quality of education matters, does religious and political beliefs play a role and also whether the level of public safety affect their behavior. As we see from Table 8, indeed some of these community characteristics affect donation behavior given the significant effects of many of the donor characteristics. We also observe considerable differences across the two donation programs. Since the intercepts for each of these variables exhibit the average effect of the donor characteristics, everything else remaining same, we will specifically focus on these intercepts, though the effects of these demographics on the time-varying covariates are also important. Let us now look at these results.

For gender, with a positive and significant effect on the intercept for the donation incidence model in case of Program 2, we argue that male donors are more likely to make a contribution for this cause (note that gender takes a value of 1 for male donors and 0 otherwise). However, gender does not play a role in the decision to make a contribution to Program 1. Neither does it affect the amount of contributions. However, gender seems to affect the responsiveness of donors to the duration from last contribution for Program 1 and the duration from solicitation for Program 2. As we look at the other individual level donor characteristics, namely the number of lifetime solicitations, we find that the overall solicitation decisions of the non-profit charity organization affect both the incidence and amount decisions for Program 1. With positive and significant coefficients for both donation incidence and donation amount, we argue that for this program, multiple solicitations not only increase the likelihood of making a contribution, but also increase the amount. The fatigue effect is absent for this program. Obviously we see lifetime solicitations affecting the responsiveness to duration from last contribution.

Now we look at the factors which explain donor heterogeneity at the community level. First we investigate the community factors which were aggregated at the ZIP Code. Median age of respondents within a ZIP Code negatively affects donation amount for Program 1 as well as the donation incidence for Program 2. This implies communities with a relatively older group of donors are less likely to contribute to Program 2 and lower amounts to Program 1. Race of donors does not affect donation behavior directly. However, for Program 1 race affects the responsiveness to duration from contribution for donation incidence and the responsiveness to duration from solicitation for donation amount. Education level of donors within a neighborhood exhibits an interesting effect on the donation pattern for Program 2. Both the incidence of donations as well as the amount is negatively affected by education, which suggests that the most valuable donors for this program are less educated, even though they are generous in terms of their frequency and amount of contributions. It might be true that in the state of Texas well to do donors are relatively less educated through they are willing to make higher donations, sometimes to attract federal tax deductions. Households with bigger sizes are less likely to contribute to Program 2 due to lower funds available. Interestingly though, when these households are families, they tend to have a higher probability of contributing to this program because they can understand the need for the causes. Household size and the percentage of households with families also influence the responsiveness to duration from contribution. To understand the effects of financial strength of households within neighborhoods we used the median income and median home values, while to measure financial stability we used the number of tradelines. Income levels negatively affect the incidence of contributions for Program 1, but home values positively affect both the incidence and amount of contribution for this program. Intuitively, this might be explained as follows. Donors with higher incomes in our dataset might actually be the ‘nouveau’ rich people who spend the newly acquired wealth on conspicuous purposes and hence contribute lower to this cause. The traditionally wealthy people are more sympathetic to the philanthropic causes and hence not only have a higher probability of contributing, but also contribute higher amounts. Donors with higher tradelines also contribute more frequently to Program 1. Even though we expected that religion to play an important role in the behavioral responses of donors, the results gives us contrary findings. For either program, we see no effects of the number of churches (both Protestants and non-Protestants) on the donation incidence and amount, Due to privacy concerns, we did not have any information on the true nature of the programs. But given that religion did not have any effect on donation behavior, we believe that the causes behind these programs were not connected to religious issues. Even though we earlier saw that the level of education seem to matter for the donation behavior towards Program 2, the quality of education does not have any effects.

Now as we focus on the county level community characteristics, we see some further interesting results. Political affiliations, measured in terms of the percentage of Republican voters affect the responsiveness to duration from solicitation for the donation amounts of Programs 1 and to a certain extent the incidence of contributions to Program 2. The final community variable, the per capita crime rate within a county, have a strong positive effect on the donation incidence behavior for Program 1 and also have effect on the responsive to duration from contribution. Communities with higher crime rates have a higher probability of contributing to Program 1. This makes us believe that Program 1 might be for a cause which actually focuses on the improvement of the quality community living.

Even though we explained some of the effects of individual and community level donor characteristics on the behavioral response of donors, overall we emphasize that there is a strong association between community characteristics and donation behavior. Since we not only see that these donor characteristics on average affect the donation incidence and amounts, but also critically influence the time-dependence of donation behavior in terms of the duration variables and past donation amounts, we argue that from a managerial point of view, it is important for non-profit charitable organizations to focus on these factors. Understanding donation behavior just in terms of past behaviors might be inadequate. Nonprofit organizations should concentrate on characterizing the donors with respect to their individual and community features. In case adequate information is not available in their donation databases regarding their donors, the non-profit charities can utilize other secondary data sources to enrich their understanding of the donors and how community characteristics affect donation behavior. This would not only help the non-profit charities to target its current donors in the future, but also identify and target potential new donors. However, we should refer back to the large scale unobserved heterogeneity present in the dataset which was discussed earlier. Even though the observed individual and community level donor characteristics do help explain some of the heterogeneity in the response pattern of donors, we should not loose sight of the potential pitfalls of the unexplained factors. Additional data on donor's attitudes and beliefs towards philanthropic causes and other psychographic factors might be helpful in explaining some of this unobserved heterogeneity.

To clearly illustrate how our specification of the hierarchical models for donation incidence and amount illustrate the role played by the community level factors, we represent the individual, ZIP Code level and county level characteristics in terms of six broad groups defining different aspects of communities (in Table 8a). These are –

1) Demographic Composition, (Proportion of male donors; Number of lifetime solicitations; Median age in years; Percentage of white; Median years of schooling; Persons per household; Percentage of households with families);

2) Financial Capability (Estimated median income; Median home value; Total number of tradelines);

3) Educational Environment (Average ratings of schools);

4) Religious Inclinations (Number of Protestant churches; Number of Non-Protestant churches);

5) Political Ideology (Percentage of Republican votes); and

6) Public Safety (Total crime per capita).

Based on these groups of donor characteristics defining communities, we show the significant effects of the variables for the donation incidence and amount models for each of the donation programs. In this representation, we clearly see the differences across the donation programs. For example, in the demographic composition of communities we see that while race (Percentage of White) matters for the donation incidence model of Program 1, it does not matter for the corresponding model for

Program 2. On the other hand, education level (Median Years of Schooling) is significant for the incidence model for Program 2, but not for Program 1. Also, 'Median Age' and race of donors does not influence the amount of contributions to Program 2, while it affects the amount of contributions to Program 1. When we look at the financial capabilities of communities, it is evident that 'Median Income' is relevant for incidence models, but not for amount models. The educational environment defined in terms of the 'Average Ratings of Schools' and public safety defined in terms of 'Total Crime per Capita' only matter for the incidence model of Program 1. Religious inclinations do not matter either for the incidence or the amount models for both programs, while political ideology defined as the 'Percentage of Republican Votes' only affects the amount model for Program 1 and the incidence model for Program 2.

Donor Segmentation and Predictive Validity

We have argued earlier that, from the perspective of the non-profit charitable organization, apart from understanding the donation behavior of its current donors, it is crucial to be able to predict the behavior of these donors in the future and also reach out to potential new donors. While for current donors, the non-profit charity organization already had available information on their past behavior, for targeting new donors, there is no historical data. That is where the importance of our donor segmentation approach becomes crucial. In that approach, we first classify donors based on the observed data about their frequency and amount of contributions. Then we characterize each of the four donor portfolios using their individual and community level characteristics. Once we are able to put a ‘face’ to the donors in each segment based on their social, economic, demographic, financial and cultural identities, from a managerial perspective is accomplishes couple of purposes. When a manager attempts to devise targeting and solicitation strategies for existing donors, he/ she does not necessarily need to look at their entire past donation behavior. Rather just by looking at the individual and community characteristics of each donor, strategies can be developed. Second, for potential donors who are not yet included in the non-profit charity organization’s database, such a segmentation approach can readily lead to an expectation regarding their behavioral responses. The donation behavior in terms of frequency and amount of donations can be extrapolated from the characterization of each donor portfolio. Such an approach for reaching both current and future donors also leads to increase in efficiency and is more actionable. When non-profit charitable organizations have large number of donors who can help achieve their philanthropic causes, individual targeting and solicitation might be very costly. Rather separate strategies can developed for each of the four cells explained in Table 1 and implemented with relative ease. Given that Segment A donors contribute less frequently and lower amounts, they can be targeted using some broad-based solicitation strategies, for example through advertisements and announcements on newspapers, televisions, internet etc. seeking donations for a cause. On the other hand, Donor Segment D is the most valuable with not only frequent contributions, but also higher amounts. This group of donors can be reached individually, carefully explaining the purpose and reach of the program and even giving them certain incentives which can improve their overall value. For the other two donor segments, namely B and C, the non-profit charity organization can adopt solicitation strategies which are a mixture of both individualized targeting as well as more aggregate targeting.

To investigate how our model is able to accurately classify donors in each of the four donor segments, we carry out a predictive validity exercise with all donors in our sample for Program 1 (note that the prediction exercise for Program 2 is not discussed since earlier we showed that there was no statistically significant difference between donors in each segment for Program 2). Since our objective is to predict donation behavior based on the donor characteristics, without explicit consideration of their past behavior, we predict the probability of contribution and the amount of contribution using the estimated intercepts and the corresponding donor characteristics. Since we had mean-centered both the time-varying covariates explaining past behavior as well as the donor-characteristics, the intercepts when multiplied with the corresponding donor characteristics have the convenient interpretation of predicted probability and amount (for the incidence and amount models respectively), when all the information from past donation behavior is kept constant at their mean levels. Based on the predictions of donation frequency and donation amount, we assign each donor to each of the four donor segments. The cutoff values of frequency and amount were determined based on the predicted mean donation frequency and amount. The means are evaluated based on the averages across all sweeps (in the estimation step for the donation incidence and amount models) for every donor and then taking the mean across all donors. The results of the prediction exercise are reported in Table 9.

In the table we report the observed number of donors in each cell, the predicted number of donors as well as the number of donors who have been classified correctly in each donor segment. The results lead us to some interesting conclusions. When we discussed the parameter estimates of the individual and community level donor characteristics to explain heterogeneity in donation behavior, we showed the importance of these variables in distinguishing between donors. However in the prediction exercise we see that we correctly assigned 51% (172 out of 335) donors in Donor Segment A (possibly the least appealing segment for the non-profit charitable organization) and 43% (71 out of 165) donors correctly in Donor Segment D (the most valuable group of donors). For segments B and C, the correct assignments are merely 13% and 18% respectively. So, overall it seems that there is some theoretical justification in using donor characteristics to identify donor portfolios in terms of donation frequency and amount (a case to the point are the high percentages for Segment A and D), the results are unsatisfactory for Segments B and C. Morrison (1969) suggests three criteria to evaluate the performance of classification schemes using discriminant analysis, namely – proportional chance criterion, maximum chance criterion and pure chance criterion. Using all three criteria, it is evident that the fraction of donors correctly classified in Segments A and D meets the performance requirements of classification suggested by Morrison[1]. Unfortunately for donor Segments B and C, such arguments cannot be made.

[1] The 'proportional chance criterion' sets a benchmark of 27% accurate classification, the 'maximum chance criterion' sets a benchmark of 33% accurate classification while the 'pure chance criterion' sets the benchmark at 25%. Since Segment A has 51% donors correctly classified and Segment D has 43% correct classification, they pass Morrison's cutoffs.

We investigated the possible reasons behind these unexpected results. First, we believe that one of the primary reasons leading to these results is the sparcity of the data. Not only do we have dataset where actual donation incidences are rare, we also see that a large proportion of donors contribute few times and also when they do contribute, the amount of contributions is considerably low. Second, since we believe that donation behavior, by its very nature, is driven to a large extent by social interactions within communities, we focused our attention to community level factors rather than individual donor characteristics. We collected the data for these community characteristics from external data sources which are inherently noisy since they had no direct association with our donation database. Third, these community characteristics were already aggregated at the ZIP Code level or the county level. So the underlying heterogeneity in donor characteristics was ignored since we assigned every donor belonging to a particular ZIP Code or county the same characteristics. We believe that if donor characteristic data was available for every donor and we were able to base our estimation on such disaggregated data, our prediction results would have improved considerably. This way we would have avoided the problem of assigning every donor in a particular community the exact same characteristic during the estimation. We could have still undertaken the predictive validity exercise using the donor segmentation approach. Fourthly, due to privacy concerns, the DMEF database does not report the true philanthropic causes behind the two donation programs we use in our research. It might well be the case that Program 1 and 2 are driven by charitable causes which do not have any direct association with social interactions within communities. Given our emphasis on community characteristics in carrying out the prediction exercises for donation frequency and amount, such lack of association between the causes and donor identities at the community level, might have lead to the disappointing results. Fifthly, we have shown earlier that even after taking into account the observed heterogeneity in terms of the community level donor characteristics, there still exists substantial degree of unobserved heterogeneity in the estimated parameters for the time-varying factors, i.e. the duration variables and the past contribution amounts. This might explain why the prediction exercise undertaken here to classify donors into the four segments using simply the donor characteristics is relatively inaccurate. Additional psychographic variables might help reduce some of the unobserved heterogeneity, thus improving the precision of the results as well as the predictions. Finally, and most importantly, it has been well documented in the marketing literature that using demographic and socio-economic information as a basis for segmentation is notoriously inaccurate because of its weak association with consumer behavior (Frank, Massy and Wind, 1972; Wedel and Kamakura, 2002). Since our objective was to predict donation behavior merely in terms of the donor characteristics without explicit consideration of past behavior to facilitate donor targeting which is managerially convenient, we carried out the predictive validity exercise accordingly. This possibly made us prone to the issues associated with the use of noisy demographic information for the segmentation of donors. Predictions might be actually better if we adopt a more structural and systematic approach for predicting donation behavior which uses some form of actual past behavioral responses or at least some additional information regarding the attitudinal and psychographic factors which influences donation behavior, as has been documented in past literature. However, we would like to emphasize here that with a more reliable donation dataset and informative donor characteristics, the segmentation approach and subsequent donor classification scheme based on predictions of donation frequency and amount would have lead to better predictions given the theoretically sound methodology employed.

Concluding Remarks

In this research, we investigated the donors' behavioral responses to direct solicitation efforts by firms in non-profit charity fundraising. We simultaneously looked at the timing of donations as well as the amount of donations in our modeling approach which provides a comprehensive tool to marketers in the non-profit charity sector to understand how contributions are made towards philanthropic causes and how past contribution patterns and solicitation strategies affect the responses. In order to explain the prevailing wide-spread heterogeneity in donation patterns, we implemented a hierarchical model formulation where the responsiveness of donors with respect to past donation behavior (driven by solicitation effort of the non-profit charities) are dependent on donor characteristics. We stress the importance of community level donor characteristics in philanthropic donations since donations are often the result of social decision-making rather than individual decision making. Finally, we provide a managerially relevant methodology to create donor portfolios based on their donation frequency and amount. However, this methodology is not driven by past behavior of donors, but rather by the characterization of donors using demographic and socio-economic identifiers.

The estimation results provide evidence of duration dependence of donation behavior, more specifically the role of duration from previous contribution and the duration from past solicitation. The impact of past donation behavior is further emphasized by the significant effect of previous donation amount on both the incidence as well as the amount of donation. We were also able to show that the individual and community level donor characteristics matter in understanding the observed heterogeneity in donations. However, due to limitations in the dataset and somewhat restrictive nature of the methodology used, we got unsatisfactory results in our predictive validity exercise. It is important to note here that the relatively greater accuracy in assigning donors to Donor Segments A and D illustrate the potential of the methodology under suitable data availability. We also point out that the individual and community-level donor characteristics which we use to explain some of the differences in donation behaviors turns out to be insufficient to explain all the heterogeneity. Significant unobserved heterogeneity prevails, as is shown using the variance-covariance matrices of the parameters.

Accordingly, the natural future research agenda involves acquiring donation databases which are less sparse with respect to actual donation events, thus avoiding some of the idiosyncrasies involved in the current dataset. Also, as mentioned earlier, availability of disaggregated donor characteristics is expected to substantially improve the predictions. Also improving the predictions can be accomplished by using additional information regarding donors past contribution behavior (in case of new donors, possibly from other non-profit charity organizations to which these potential donors were already associated with) or at least their attitudes towards philanthropic causes. Finally, this research focuses on the behavior of donors in response to solicitation and targeting efforts by non-profit charitable organizations. The next obvious question is, how do the non-profit charities decide whom to target and when to target? This question becomes even more relevant given the weak prediction results we obtained regarding the donation behavior when donor characteristics are used for donor segmentation and targeting purposes. The results from the model estimation can be incorporated in a sequential decision making algorithm in order to determine the optimal mailing decisions. In addition to the actual value of each donor in terms of their contribution patterns, the non-profit charity organization can include an updating mechanism in their beliefs about the donors behavioral responses and this information learning would help them better predict the donors behavior in the future. Such an approach which constructs a multi-period decision problem for the non-profit charitable organization with a value-information tradeoff is expected to perform better than any ad-hoc approach to customer targeting. It is also likely to be more efficient compared to an algorithm which looks at a single period decision or simply maximizes the value of donations or the information content of donations individually. This would optimize the non-profit charitable organization's solicitation strategies over the donors' lifetime.

CHAPTER THREE:

SEQUENTIAL DECISION MAKING AND THE VALUE OF INFORMATION FOR OPTIMAL COUPON TARGETING IN RETAIL

Background

Direct marketing in the context of retailing is a relatively new marketing initiative. Usually retailers target customers through in-store price promotions and displays or coupons, which sometimes carry monetary value or offer bonus buys, thus effectively reducing per unit pricing of the product. The common practice in retailing is to provide coupons as inserts in newspapers and is not specific to customers' purchase patterns in a product category. Recently relatively more customer-specific coupons are increasing in usage, either as in-store coupons which can be redeemed at point of purchase, or sometimes printed during check-out (the Catalina Marketing Incorporated's Checkout Coupon system), which can be redeemed during subsequent purchase occasion. Such coupons are more customized since they are based on the basket-mix of the customers during a specific purchase occasion.

This point of purchase couponing system however does not take into account the purchase behavior of the customer in a particular product category. It only considers a specific purchase occasion. On the one hand, the customer might have purchased in the product category just as an aberration and might be unlikely to make another purchase soon. Since coupons usually have an expiration date, it might not be worthwhile providing a coupon to this customer. On the other hand, if a customer is a frequent purchaser in a product category, irrespective of the coupon provided to him/ her, a purchase is likely to be made. So mailing this group of customers a coupon might again turn out to be an inefficient proposition. However, there is a third group of customers whose decisions to make a purchase in the product category are often driven by the availability of coupons as well as the face value of those coupons. If the retailer can devise a customized couponing mechanism whereby the face values of coupons are determined based on the customers' past purchasing behavior as well as their expected future behavior, they will be able to attract a large section of these undecided group of customers. The retailer's decision is this context is whether to mail a coupon to a customer, at what time and if so what should be the corresponding face value.

Another issue faced by retailers when devising targeting strategies with coupons is – in the event that a decision is made to mail a coupon to a customer, which brand's coupon is most appropriate, since the customers' brand preference will play a significant role in determining the purchase behavior of the customer in that product category and also how they redeem their coupons. Since coupons at the brand level are usually provided by the manufacturers, when the retailer chooses to mail customized coupons to its customers, the decision which they need to make is which brand's coupon should be used for the targeting after collaborating with the manufacturer to determine the face values of the individually customize coupons. The current prevailing practice in retailing is to send coupons at the category level. However, brand level couponing and its effects on purchase behavior as well optimal targeting using customized coupon face values is getting some traction in retailing. Brand level couponing facilitates brand switching, thus the retailers can potentially use the optimal targeting mechanism to increase the share of high margin brands at the cost of low margin brands which maximized category profits. Retailers can set the terms of their contracts with the manufacturers in such a way that customization regarding face values and the timing of coupon mailing can be coordinated accordingly.

In this research, we focus on direct marketing initiatives in the context of retail sector by first constructing a customer response model of purchase within a product category, where purchase decisions are driven not only by the prices and other in-store promotional strategies of retail marketers for brands within that category, but also by the individual customer – specific brand-level couponing schemes implemented as part of direct marketing initiatives. In marketing, one of the few research works which looks at direct marketing in the context of the retail sector is by Rossi, McCulloch and Allenby (1996) where the authors focus on a household-specific target couponing problem with the customizations done based on the information set available. Their customer response model is built as a discrete choice formulation with a flexible heterogeneity structure accommodating both observed and unobserved differences in customer behavior. The information sets in this set-up are obtained from current and past purchase history data as well as customers' demographic characteristics. They subsequently determine an optimal coupon targeting strategy by optimizing over the face value of targeted coupons such that the incremental revenue due to coupon discounts are maximized.

We augment their customer response model by introducing duration dependence of brand-level choice behavior of customers in addition to incorporating the effects of pricing and other in-store promotions. We also include the role of past targeting actions of retailers, specifically the effects of coupon mailing in the purchase decisions within a product category. Further, even though our retail targeting strategy involves determining optimal coupon face values (similar to the work of Rossi, McCulloch and Allenby), in the development of the optimal coupon targeting problem we address two critical problems which has not been considered in the literature – First, in addition to using the incremental value of coupons in retail purchases as the return from customized targeting, we use a augmented utility function of the retailer which includes the value of information, thus providing the retailer market intelligence via information gathering during the decision making process; Second, we construct a sequential decision making algorithm where the retailer's current decisions are not only based on the past response behavior of the customers, but also the future expected behavior. This essentially translates into a updating of beliefs over the decision time horizon which is expected to increase the value of the customers over their lifetime.

The rest of this chapter is organized as follows. First, we discuss the literature in customer response models in the context of the retail sector and optimal targeting. We focus on response models which has applications in the area of direct marketing and how researchers have conceived optimal targeting approaches using decision theoretic methodologies. In the next section, we intuitively explain the two stages of my model – first, the customer response model of brand choice within a category affected by targeted coupons, pricing and in-store promotions and second, the optimal coupon targeting and the value of information. The following section focuses on the mathematical details of the customer response model as well as the optimization algorithm. In the customer response model, we explain the steps involved in formulating the response model within the competing risk hazard model framework. In the discussion of the optimal targeting algorithm, we adopt a two-step approach. First we introduce the information gain in the retailer's utility function to explain its relevance in improving the targeting. Subsequently, we construct a sequential decision problem where the retailer updates his/ her beliefs based on the expected redemption patterns of the customers. We end this section by discussing the details of the optimization algorithm and how each of the potential computational issues are addressed. We end the chapter by providing some concluding remarks regarding the contribution of this research and potential shortcomings.

Literature

Customer choice models in retailing have been a prolific area of research for marketing scholars. Works by Kamakura and Russell (1989), Bucklin and Gupta (1992), Jain et al. (1994), Fader and Hardie (1996) are some of the earlier and prominent works in this area. The main focus of these papers were to model purchase behavior in a retail setting where customer response behavior was captured in terms of purchase incidence, brand choice or even choice behavior between SKUs. These authors used alternative assumptions regarding the heterogeneity of customer behavior ranging from heterogeneity with discrete mass points to continuous heterogeneity as well as mixture distributions defined over a continuous space.

Though direct marketing and more specifically formulation of individualized targeting strategies in the context of retailing has not been explored in the marketing literature, customer response models with focus on direct marketing initiatives has been studied in other contexts. The pioneering work by Schmittlein and Peterson (1994) model the expected response behavior of active customers using data on frequency, timing and dollar value of past transactions. Other notable works in this area include Basu et al. (1995) who propose the ‘heterogeneous starting point model’ to model response pattern to direct marketing campaign, Allenby et al. (1999) who model customer interpurchase times allowing for both temporal and cross-sectional heterogeneity and Baesens et al. (2002) who model learning from customer repeat purchases using Bayesian neural network approach. Gonul and Ter Hofstede (2006) who propose the estimation of a response model of order timing and order volume decisions of catalog customers, also implements a Bayesian decision theoretic approach to derive optimal mailing strategies for catalog customers in their article.

Such optimal mailing strategies has been also been explored in the direct marketing literature frequently in case of catalog marketing, but seem to lack an application in retailing. Bult and Wansbeek (1995) propose a method for selecting targets from mailing list by equating marginal returns to marginal costs, Bitran and Mondschein

(1996) study optimal mailing policies in catalog sales industry in a stochastic environment with random customer responses and dynamic evolution of mailing list whereas Gonul and Shi (1998) suggest a dynamic programming model for optimal mailing by maximizing utility for customer and profits for direct marketers. Additionally, Piersma and Jonker (2004) develop a model to determine optimal mailing frequency to achieve a profitable long-term relationship, Elsner et al. (2004) address the optimal decisions involving the questions of when to mail and whom to mail from a medium-term perspective instead of a completely myopic decision process and Simester et al. (2006) take into consideration the dynamic implications of optimal mailing decisions in order to understand their long-term implications.

In this research, we propose a Bayesian decision theoretic approach for direct marketing in the retail sector. Given the limited research on direct marketing in the retail sector, our objective is to explore how customized coupon targeting strategies can provide retailers a powerful tool to reach each customer individually using information regarding their past purchase behavior. The algorithm we propose to determine the optimal decisions regarding the face value of coupons, the customers to target and the timing of this targeting improves on the existing understanding in Bayesian decision theory in two ways – first, we stress the importance of gaining information during targeting in addition to maximizing revenue; second we focus on a decision making process which leads to updating of beliefs over time in a sequential optimization framework.

Model Outline

<u>Customer Response Model – Category Purchase</u>

We build our customer response modeling framework primarily based on the approach used by Rossi, McCulloch and Allenby (RMA henceforth), who model category purchase as a brand choice decision affected by coupon redemption and other marketing factors with a continuous heterogeneity distribution assumption for the customer level parameters. Our model has the following notable features, some of which follow form RMA, and some which we introduce to highlight important aspects of customer brand choice and retail targeting –

First, RMA build their target couponing problem using purchase history data of alternative brand choices within a product category. However, in order to maintain the basic structure of a discrete choice model, they do not consider how past couponing behavior, both in terms of targeting by the retailer and redemption by the customer, affects brand purchase within that category. The rationale behind this was their ultimate aim to understand the role of alternative information sets (baseline, demographic, choice-based and full information sets) in making the targeting decisions, where they showed that even minimal information available regarding purchase history immensely improves the targeting strategies. We modify their approach by introducing duration dependence in brand choice where past behavior, both with regards to purchase and coupon redemption influences decisions. This is because customers who have made a purchase recently are unlikely to make another purchase and make the decision regarding redeeming of the coupon soon because of the costs associated with maintaining a stock of the product in question. On the other hand, if the coupon is really attractive to the customer because of its face value, the trade-off between stockpiling costs and savings on the product becomes critical. Thus, while RMA use a discrete choice model to formulate the purchase behavior of customers within a product category, where coupon redemption essentially translates into an increase in the category purchase probability, we employ a hazard model formation with duration dependence, thus allowing for past purchase and coupon redemption to affect current decisions. We consider a relatively simple structure to model customer-level heterogeneity and use Normal distributional assumptions on the response parameters. Alternatively, we could have uses a Normal continuous mixture model such that the heterogeneity in the propensity to redeem coupons while making a brand choice decision is captured in terms of the continuous densities. Though this is important, we believe that even with the simple heterogeneity structure the differences in the propensities can be captured. It is important to note that, even though we model purchase probability in terms of the hazard model, we do not consider the monetary value of purchases within each purchase occasion, primarily because in a retail setting usually the brands within a product category are priced very similarly and accordingly the amount spend during each purchase occasion in that product category is unlikely to vary significantly.

Second, we believe that the past targeting by the retailers using coupons for the various brands within the product category affects the purchase behavior and possibly coupon redemption as well. Accordingly, we introduce indicator variables for each brand in our model denoting the incidence of targeting via coupons for each customer during the observation period, thus capturing the history of targeting via coupons to each customer. This is important because coupons send by the retailer not only creates awareness about the deals available for the alternative brands, but it might also enable the customer to evaluate their relative face values. This becomes even more critical under the assumption that the retailer can send coupons for only one brand at any given time period because of the competitive nature of the brands within the category and also customers can redeem only one coupon within the available coupons due to possible budgetary constraints as well as issues of stockpiling costs.

Third, in retail setting marketers are usually faced with the challenge with regards to the choice of coupons to send. Based on the contractual negotiations with the manufacturers, retailer usually can choose between alternative coupons from the different brands within a category at any point in time. So, the decision problem of the retailer not only involves which customer to send a coupon and when, but also they need to decide on which coupon to send. Accordingly, in our modeling framework we choose to use a competing risk hazard model where the decision to redeem coupon from one brand within the product category essentially precludes the redemptions for the other coupons. Subsequently brand choice is also a competing risk hazard process. This essentially means that once any of the brands is purchased, the duration from the time of last purchase is reset for all the brands within the product category. To keep things simple we assume that the hazards are uncorrelated, thus making the hazard of purchase decisions for each brand independent. Also, it is important to note that even though the customer can make a brand purchase at any point in time, be it with or without coupon redemption, we observe the data only in discrete intervals. Following the literature on hazard models with interval-censored data, we employ a complimentary log-log specification for the hazard function (details explained in the next section).

Fourth, following the framework used by RMA, we believe that during the purchase of a particular brand within a product category, not only are own prices important, but also cross prices play a significant role. For example, a customer who might have been loyal to Brand A at a given price level, might be interested in switching to Brand B if after the redemption of a coupon for Brand B, prices for Brand B are lower than that of Brand A. Such brand switching behavior is usually the primary rationale for the use of brand level coupons, and more specifically customized couponing schemes which can effectively capture the loyalty of customers by tracking their previous purchase history.

Finally, we introduce dummy variables which indicate the presence of in-store promotions during the observation period for each customer. These promotions not only include in-store displays promoting a brand in terms of in-store coupons, samples etc., but also incorporates in-store advertisements which illustrate the features of each brand. RMA considered in-store displays and feature advertisements as separate factors, but we chose to combine all possible store-level marketing efforts in order to limit the dimensionality of our model, while at the same time capturing the essential components driving retail purchase behavior. We believe that in constructing a customer response model which formulates purchase behavior within a category, these factors are extremely important. Not only are these in-store promotions likely to increase the purchase probabilities independent of the availability of customized targeted coupons, but also might serve as an alternative method of targeting customers when costs of targeting individual customers are higher. On the one hand, the influence of in-store promotions might mitigate the effectiveness of customized coupons to a certain extent, but on the other they might also reinforce the purchase intentions in the presence of coupons.

Optimal Couponing – Value of Information

Our approach in devising the optimal couponing problem, however, differs significantly from the proposed methodology of RMA. They use a very stylized couponing problem with a exclusive focus on the revenue side of couponing without explicitly considering the costs associated with issuing and redeeming coupons. More importantly, RMA models couponing as temporary price cuts where customers receive targeted coupons and retain them for possible redemption during subsequent purchase occasions. Thus the value of targeted couponing is to facilitate brand switching within a product category motivated by the differential margins for the retailer across the alternative brands. Accordingly, we believe that RMA's approach essentially considers the issuing of coupons as a one-time strategy which only takes into account the previous purchase occasion and not the entire history of customers' purchases within the product category, while at the same time ignoring the expected purchase behavior of the customer or their redemption of the coupons. The decision only involves the choice of the face value of each customized coupon, which inherently considers the scenario where no coupon is send, by allowing for zero coupon values.

In devising the optimal couponing problem, we address two important questions which are pertinent for the direct marketers towards formulating an optimal targeting strategy using coupons –

1) What should be the objective function of the retailer when trying to make an optimal targeting decision using coupons? The obvious objective is to maximize the expected revenue stream from their customers based on their purchases. However such a utility function focuses only on the scenario where the customer actually responds to the marketing initiatives of the retailer including the targeted coupons, by choosing one of the brands. Given the low response rates in direct marketing as well as the customer heterogeneity involved, there might be valuable information regarding the customer's behavior even when the customer does not respond by making a purchase or decide not to redeem a coupon. So the question is – as a direct marketer, do optimal targeting decisions just involve optimizing over expected revenue driven by the customer's brand purchase behavior or is there an additional value in learning about customer behavior by continuously updating the information from customer responses as well as non-responses? This information update provides marketers with market intelligence since during the optimal decision making they are not only maximizing their payoffs but also improving the precision of their decisions.

2) What should be the time horizon over which the direct marketer makes the targeting decisions in retailing? In most cases the decision making process is myopic and the targeting strategy just involves maximizing the marketer's utility one period at a time. Alternatively the marketer tries to optimize the expected revenue stream from its customers over a finite time horizon (which is usually the lifetime of the customer). This targeting decision implicitly assumes that there is no evolution in the customer's behavior in the future periods. The customer is expected to respond in the same was over his entire lifetime and make consistent purchase decisions in the product category. However, that seems to be a debatable assumption. Customers evolve continuously in their response behavior with respect to brand purchase behavior in a retail setting and so it is important for the direct marketer to also learn from the customer's behavior. This makes the decision making process is sequential since the expected decisions of the direct marketer in the future periods affect the marketer's decisions in the current period in addition to their past behavioral patterns.

In this research we address these questions individually and then simultaneously.

Step 1: First, we construct an algorithm for the optimal targeting decision in direct marketing in the retail setting using an utility function specification for the direct marketer which not only incorporates the expected revenue from the customers based on their brand choice and making a purchase in the product category (influenced by targeted coupons, pricing and other in-store promotions), but also an additional term which captures the learning through information update. We capture this information update in terms of the Shannon Information Measure which is defined as the ratio of the posterior distribution of the parameters given the data in the current period and the prior distribution of the parameters, which are essentially the posterior distribution of parameters in the previous period. This entropy measure captures the uncertainty associated with the parameter vector and with the inclusion of this term we make sure that the information content of each stage of the decision making process is captured. In other words, this measure captures improvement in precision of the parameters in every consecutive period of decision making, thus implicitly improving the precision of the targeting mechanism for coupons. In order to update the prior parameter estimates as a new wave of data becomes available, we use a particle filtering algorithm with weights proportional to the likelihood and then make draws from the posteriors to calculate the Shannon Information Measure. Since the decision making is one-period ahead, we simulate pseudo observations for the subsequent period by drawing parameters from the posterior distribution. Note that, here we look at a scenario where the decision maker is myopic and only makes decisions one period ahead (albeit using a modified utility function), though this assumption is relaxed in the next step.

Step 2: In the second step of this targeting approach for couponing, we propose an algorithm which considers the direct marketer's optimal decision making process regarding the issuing of coupons as a multi-period problem. However, instead of assuming that the customer's response behavior to coupons does not evolve over time, we take into account the possible changes in customer response patterns and thus construct the problem as a sequential decision problem from the point of view of the direct marketer. In this approach, we initially assume that the utility function specification just takes into account the expected revenue stream from the customers, though it is subsequently relaxed to include incorporate the modified utility function explained before. Under such sequential decision problems, the decision of sending coupons is dynamic and so the solution algorithm involves solving a backward induction problem. However, this backward induction process introduces two computational issues. First, we need to take into account all the possible decision paths as we move back from the last period to the current period and calculate the expected utilities in each of those paths. Second, in each of these decision paths we need to optimize over the possible decisions to arrive at the best targeting strategy using the customized coupons, thus requiring numerous optimization routines at each step of the way. As a solution to these problems we use forward simulation in conjunction with a constrained action space. In forward simulation we generate multiple simulated experiments and then start the backward induction algorithm from the last period by optimizing over the expected utilities in that period. As we come back one period at a time, we consider only those paths from the simulated experiments which were optimal in the next period as the relevant decision paths and then optimize over the expected utilities in those paths. Also, in order to reduce the dimensionality of the information available in all the possible paths, we summarize this information with two sufficient statistics, namely the mean and the variance of the value/ utility function, and this eases the computational burden of calculating the expected utilities corresponding to each decision path.

Finally, we combine Step 1 and Step 2 and propose an algorithm which not only incorporates a utility function specification for the direct marketer as convex combination of expected revenue and expected information gain, but also position the optimal targeting decisions as sequential, thus introducing endogeneity in the decision making in every period. It is expected that in the initial periods when the direct marketer undertakes this sequential targeting strategy, he might need to sacrifice some of the expected incremental revenue stream from its customers in order to gather more information about the customer. But as the decision time horizon gets longer, the learning through information update reaches a saturation point (no additional information improves the precision of the targeting decision), the relative value of information goes down while the expected revenue from the customers keeps increasing with better understanding of the customer's behavior. This is illustrated in Figure 3. As illustrated in the figure, we expect that if we take into account the update of information obtained from the purchase behavior of customers, the expected flow of incremental revenue over the decision making time horizon would be an upward sloping curve. Revenue stream in the initial periods of the sequential decision would be lower compared to the curve without information update because the retailer at that point would be gathering more information for a better targeting using the coupons. Over the customer's lifetime, the revenue stream increases at a greater rate than the value of information, and eventually flattens out as no additional information needs to be collected. On the other hand, if we do not consider the possibility of information update, we would also see an upward sloping curve, which obviously gives us higher revenue in the earlier time periods given that the retailer targets only targets those customers with a coupon who provide the maximum incremental revenue. However, this does not consider the evolution of customer behavior, and eventually with no learning, the retailer gradually fails to gain additional revenue because the customer’s brand purchases has reached a saturation point. Such stagnation explains the curve flattening out much earlier than the curve with information update. The framework explained above is essentially a natural experimentation method. The retailer in the initial period experiments with which customers to target, when to target and with which brand’s coupon. Over time, the information gain makes the retailer experienced in their targeting, and he/ she can target specific coupons to specific customers.

Mathematical Outline of Model

Customer Response Model using Competing Risk Hazard Framework

Let us first consider the customer response behavior in making a brand choice within a product category. Understanding of the response behavior is crucial for the retailers before considering the optimal targeting strategies which are obviously based on the response patterns. In our setup, at any time period, the customer decides whether or not to purchase a particular brand within a product category. During each purchase occasion the customer also decides whether not to redeem a coupon for that brand in question.

With this focus, we formalize the brand choice problem as a competing risk discrete-time proportional hazard model, because as explained earlier the purchase behavior is duration dependent and also when making a choice the customer can choose only one of the brands from among many in the product category, thus making brand choices interdependent. We index customers by $i = 1, 2, \ldots n$ who can purchase from brands $j = 1, 2, \ldots J$ in a particular product category in periods $t = 1, 2, \ldots T$.

Let $y_{ijt} = 1$ if customer *i* purchases brand *j* at period *t*

= 0 otherwise

For any customer *i*, the hazard of purchasing brand *j* at time period *t* is given by –

$$(1) \quad h(d_{ijt}^{c}, X_{ijt}) = \lim_{\Delta t \to 0} \frac{P(d_{ijt}^{c} \le T_{ijj^{/}} < d_{ijt}^{c} + \Delta d_{ij}^{c} \mid T_{ijj^{/}} > d_{ijt}^{c})}{\Delta d_{ij}^{c}} = h_0(d_{ijt}^{c}) \exp(\beta_{ij}^{/} X_{ijt})$$

where $T_{ijj^{/}}$ is the interval between purchases in the product category for brand *j* and $j^{/}$ with $j, j^{/}$=1,…,J .

This proportional hazard function specification above has two components. The first part $h_0(d^{c})$ is a time-varying component which captures the dependence of brand choice on previous choices made in terms of the durations between purchases (d_{ijt}^{c}) and is the baseline hazard. The second part $\exp(\beta_{ij}^{/} X_{ijt})$ corresponds to all other time-varying and non-time-varying factors which might influence the hazard of making a purchase of brand *j* in the product category.

For each brand *j*, the likelihood for customer *i* making a purchase is subsequently written as - $\left(h(d_{ijt}^{c}, X_{ijt})\right)^{y_{ijt}} \left(1-h(d_{ijt}^{c}, X_{ijt})\right)^{1-y_{ijt}}$ with the overall likelihood of the data given by - $\prod_{\forall i}\prod_{\forall t}\left(h(d_{ijt}^{c}, X_{ijt})\right)^{y_{ijt}} \left(1-h(d_{ijt}^{c}, X_{ijt})\right)^{1-y_{ijt}}$

Since our brand purchase data is interval censored, when constructing the likelihood and considering the probability of observing an exit to a specific destination state (an actual purchase event) in a given interval, we not only need to take into account the fact that there was an exit to that destination, but also that the exit occurred before an exit to other potential destinations. By making use of the relationship between the overall discrete interval-censored hazard and the destination-specific interval hazards, and the relationship between the discrete interval hazards and the underlying continuous hazards, it can be showed that –

(2) $$h(d_{it}^{c}, X_{it}) \approx \sum_{j=1}^{J} h(d_{ijt}^{c}, X_{ijt})$$

In the above expression, it is seen that the overall interval hazard is only approximately equal to the sum of the destination-specific hazards.

The corresponding survivor function at time *t* for brand *j* is represented as -

(3) $$S(d_{ijt}^{c}, X_{ijt}) = \exp\left[-\int_{0}^{d_{ijt}^{c}} h_0(u_{ijt}^{c})\exp(\beta_{ij}^{/} X_{ijt})du_{ijt}^{c}\right] = \exp[-\exp(\beta_{ij}^{/} X_{ijt})H(d_{ijt}^{c})]$$

$$\text{where} \quad H(d_{ijt}^{c}) = \int_{0}^{d_{ijt}^{c}} h_0(u_{ijt}^{c})du_{ijt}^{c}$$

Following the same line of reasoning as in the case of the hazard function, the survivor function for exiting to any destination can be written in terms of the survivor functions for exit to each destinations as –

$$S(d_{it}^{c}, X_{it}) = \prod_{j=1}^{J} S(d_{ijt}^{c}, X_{ijt}) \tag{4}$$

Before proceeding to write-out the likelihood of the data generating process, it is important to make a crucial assumption which greatly simplifies the estimation of a competing risk discrete hazard model. First suggested by Narendrenathan and Stewart (1993), it essentially implies that transitions between states can only occur at the boundaries of the intervals. Using this assumption, the likelihood of purchase for brand *j* can simply be written as –

$$\Pr(y_{ijt} = 1) = \left[\frac{h(d_{ijt}^{c}, X_{ijt})}{1 - h(d_{ijt}^{c}, X_{ijt})} \right] \prod_{j=1}^{J} S(d_{ijt}^{c}, X_{ijt}) \tag{5}$$

Defining the destination-specific (purchase event) censoring indicators as δ^{j} , the overall likelihood for customer *i* at any time period *t* is given by (note that this is the probability that customer effectively makes a purchase in the product category) -

$$\Pr(Purchase) = \prod_{j=1}^{J} \left[\frac{h(d_{ijt}^{c}, X_{ijt})}{1 - h(d_{ijt}^{c}, X_{ijt})} \right]^{\delta^{j}} S(d_{ijt}^{c}, X_{ijt}) \tag{6}$$

This simplified expression implies that when estimating the overall competing risk interval-censored discrete hazard model, one can estimate separate destination specific hazard models after defining suitable destination specific censoring variables.

Due to this interval-censoring of hazard data for the purchase of brand *j* (which is affected by the prices for all brands, in-store promotions, the coupon targeting by the retailer for all brands as well as the coupon face values when redeemed), the discretization of the hazard function specification gives us –

$$\begin{aligned} h(d_{ijt}^c, X_{ijt}) &\approx P(T_{ij} = d_{ijt}^c \mid T_{ij} > d_{ijt}^c - 1) = \frac{S(d_{ijt}^c - 1, X_{ijt}) - S(d_{ijt}^c, X_{ijt})}{S(d_{ijt}^c - 1, X_{ijt})} \\ &= 1 - \exp\left[-\exp(\beta_{ij}{}' X_{ijt})\left(H(d_{ijt}^c - 1) - H(d_{ijt}^c)\right)\right] \end{aligned} \tag{7}$$

Note that, this hazard function is the conditional probability of observing a purchase at time period *t* given no brand purchases have been made till period *(t-1)*.

Equation (7) can be rewritten as –

$\log\left(1 - h(d_{ijt}^c, X_{ijt})\right) = \exp(\beta_{ij}{}' X_{ijt})\left(H(d_{ijt}^c - 1) - H(d_{ijt}^c)\right)$, which gives us –

$$\log\left(-\log(1 - h(d_{ijt}^c, X_{ijt}))\right) = \beta_{ij}{}' X_{ijt} + \log(H(d_{ijt}^c) - H(d_{ijt}^c - 1)) \tag{8}$$

The baseline hazard function is then -

$$\log\left[-\log(1 - h_0(d_{ijt}^c))\right] = \log(H(d_{ijt}^c) - H(d_{ijt}^c - 1)) = \log\left[\int_{t-1}^{d_{ijt}^c} h_0(u_{ijt}^c)\, du_{ijt}^c\right] = \gamma_{ijt} \tag{9}$$

Substituting terms in equation (8), we get: $\log(-\log[1-h(d_{ijt}^{c}, X_{ijt})]) = \beta_{ij}^{/} X_{ijt} + \gamma_{ijt}$ which leads us to the final form of the hazard function as –

$$h(d_{ijt}^{c}, X_{ijt}) = 1-\exp[-\exp(\beta_{ij}^{/} X_{ijt} + \gamma_{ijt})] \tag{10}$$

Equation (10) specifies the final form of the hazard function (a complimentary log-log model) corresponding to the purchase of brand *j* in the product category. The baseline hazard function in the above specification captures the effect of the duration from previous purchases. Given the mean-centering of the duration variable, we not only include an intercept, but also to capture possible non-linear effects of the duration variable on the brand purchase behavior, both linear and log terms are included. Accordingly, the baseline hazard function is –

$$\gamma_{ijt} = \alpha_{ij0} + \alpha_{ij1} d_{ijt}^{c} + \alpha_{ij2} \log(d_{ijt}^{c}) \tag{11}$$

Here α_{ij0} is the intercept, α_{ij1} is the linear effects of the duration from past purchase occasion and α_{ij2} is the corresponding non-linear effect. These linear and non-linear terms together capture any possible non-monotonicity in the effect of the duration variable. Based on our discussion earlier, additional factors which influence the hazard of brand choice by the customers are included in $\exp(\beta_{ij}^{/} X_{ijt})$. We define the linear predictor $\beta_{ij}^{/} X_{ijt}$ for each customer *i* corresponding to brand *j* as –

$$\beta_{ij}^{/} X_{ijt} = \sum_{j^{/}=1}^{J} \phi_{ijj^{/}} (p_{ij^{/}t} - f_{ij^{/}t}) + \sum_{j^{/}=1}^{J} \lambda_{ijj^{/}} m_{j^{/}t} + \sum_{j^{/}=1}^{J} \mu_{ijj^{/}} q_{ij^{/}t} \tag{12}$$

for all j=1,…,J .

where $p_{ijt}(p_{ij't})$ is the price which customer i pays for brand $j(j')$ at time t, $f_{ijt}(f_{ij't})$ is the face value of the coupon being offered for brand $j(j')$ at time t by the retailer, $m_{jt}(m_{j't})$ denotes the dummy variable indicating whether there was any form of in-store promotion (display, feature advertisements etc.) available for brand $j(j')$ at time period t and $q_{ijt}(q_{ij't})$ is the dummy variable denoting whether the retailer targeted customer i at time period t with a customized coupon for brand $j(j')$.

Interpreting the coefficients in equation (12) gives an insight into how the retailer's targeting using coupons and the face values of those coupons affect the purchase probabilities. For the moment assume that since we are only concerned about the brand choice for j, the prices, in-store promotions as well as the coupon targeting by retailers and corresponding coupon face values $\{p_{ij't}, f_{ij't}, m_{ij'}, q_{ij't}, j \neq j'\}$ for competing brands are kept unchanged. In this case, Table 10 provides a nice interpretation of the parameters $\{\phi_{ij}, \lambda_{ij}, \mu_{ij}\}$. In Table 10, we have coupon offered (No/ Yes) on the vertical axis and coupon redeemed on the horizontal axis. It is important to note here that offering a coupon for a brand j by the retailer does not necessarily translate into redemption of the coupon or the purchase of brand j. It depends on the sensitivity of the customers to price, the advertising awareness created by the coupon, the face value of the coupon as well as the in-store non-price promotions. Assuming that $m_{ij} = 0$, in the case where no coupon is offered $(q_{ijt} = 0)$ and subsequently nothing is redeemed during purchase $(f_{ijt} = 0)$, the effect on the brand purchase probability for j is given by $\phi_{ij} p_{ijt}$; if a coupon is offered $(q_{ijt} = 1)$, but not subsequently not redeemed during purchase $(f_{ijt} = 0)$ (either because the customer purchased a different brand or decided not to redeem even though he/ she had the coupon available), the effect on the brand purchase probability for j is given by $(\phi_{ij} p_{ijt} + \mu_{ij})$. So here effectively μ_{ij} is the effect on the brand purchase probability which is driven by the fact that the retailer had sent a coupon. This is not contingent on actual redemption. Essentially we can view this as an advertising effect of the coupon, where the coupon itself creates awareness in the customer's mind regarding the availability of the brand in question. Of course, the case where there is no coupon offered $(q_{ijt} = 0)$, there is obviously no possibility of redemption. Finally, when the retailer chooses to offer a coupon for brand j $(q_{ijt} = 1)$ and the customer also chooses to redeem the coupon during the purchase of brand j $(f_{ijt} = 1)$, we see that the effect on the purchase probability is given by $(\phi_{ij} p_{ijt} + \mu_{ij} - \phi_{ij} f_{ijt})$. So $\phi_{ij} f_{ijt}$ is the effect on the purchase probability based on the face-value of the coupon over and above the advertising effect. \

<u>Specification of Heterogeneity</u>

The response parameters for the model of purchase behavior in the product category (expressed in terms of equations (11) and (12)) are customer specific. As we have explained earlier, the heterogeneity in the behavioral responses of customers are modeled using Normal distributional assumptions. Accordingly, for the brand choice model, we incorporate the heterogeneity in customer's behavior using the following specification –

(13)

$$\begin{pmatrix} \alpha_{ij0} \\ \alpha_{ij1} \\ \alpha_{ij2} \\ \phi_{ij j'} \\ \lambda_{ij j'} \\ \mu_{ij j'} \end{pmatrix} \sim N\left(\begin{pmatrix} \bar{\alpha}_{j0} \\ \bar{\alpha}_{j1} \\ \bar{\alpha}_{j2} \\ \bar{\phi}_{jj'} \\ \bar{\lambda}_{jj'} \\ \bar{\mu}_{jj'} \end{pmatrix}, \Sigma \right) \quad \text{where } \phi_{ijj'} = \begin{pmatrix} \phi_{ij1} \\ \phi_{ij2} \\ \ldots \\ \phi_{ijJ} \end{pmatrix}; \lambda_{ijj'} = \begin{pmatrix} \lambda_{ij1} \\ \lambda_{ij2} \\ \ldots \\ \lambda_{ijJ} \end{pmatrix}; \mu_{ijj'} = \begin{pmatrix} \mu_{ij1} \\ \mu_{ij2} \\ \ldots \\ \mu_{ijJ} \end{pmatrix}$$

$$\text{and } \bar{\phi}_{jj'} = \begin{pmatrix} \bar{\phi}_{j1} \\ \bar{\phi}_{j2} \\ \ldots \\ \bar{\phi}_{jJ} \end{pmatrix}; \bar{\lambda}_{jj'} = \begin{pmatrix} \bar{\lambda}_{j1} \\ \bar{\lambda}_{j2} \\ \ldots \\ \bar{\lambda}_{jJ} \end{pmatrix}; \bar{\mu}_{jj'} = \begin{pmatrix} \bar{\mu}_{j1} \\ \bar{\mu}_{j2} \\ \ldots \\ \bar{\mu}_{jJ} \end{pmatrix} \text{ for } j' \neq j$$

Σ is the variance-covariance matrix for the factors included in X_{ijt} .

Estimation

The estimation for the customer brand purchase model (which effectively leads to the prediction of purchase probabilities) can be done using Markov Chain Monte Carlo (MCMC) methods. Note that the estimation is done using mean-centered data, the time-varying covariates, namely the duration from previous purchase occasion as well as the brand level prices, brand level in-store promotion, brand level history of coupon targeting and the face values of brand-specific coupons, are all mean-centered with respect to their customer-level means. In Bayesian MCMC methods, posterior distributions are approximated by sampling from the full conditional distributions where full conditionals are obtained from the prior distribution of the parameters and the observed data using Bayes rule. In the estimation of the brand purchase model, we do not have closed form representations of the full-conditional distributions for the customer-specific parameters $\{\Theta_{ij} = \{\alpha_{ij0}, \alpha_{ij1}, \alpha_{ij2}, \{\phi_{ijj'}, \lambda_{ijj'}, \mu_{ijj'}\}_{j' \neq j=1,2,...,J}\}\}$. So we use the Metropolis-Hastings algorithm with a Normal distribution for the proposal densities involved in the Metropolis-Hastings steps.

The response parameters are customer specific which allows for different response patterns across customers. While unobserved customer characteristics might affect the shape of these curves, the response parameters may also be correlated. So we assume the following distributional specifications for the customer-specific parameters which allows for the unobserved factors –

$$[\Theta_{ij} = \alpha_{ij0}, \alpha_{ij1}, \alpha_{ij2}, \{\phi_{ijj'}, \lambda_{ijj'}, \mu_{ijj'}\}_{j' \neq j=1,2,...,J}]' \sim N([(\bar{\alpha}_{j0}, \bar{\alpha}_{j1}, \bar{\alpha}_{j2}, \{\bar{\phi}_{jj'}, \bar{\lambda}_{jj'}, \bar{\mu}_{jj'}\}_{j'=1,2,...,J})]', \Sigma)$$

Note that the posterior distributions of the customer level parameters for the brand choice model are approximated using the Metropolis-Hastings algorithm.

The prior distributions for the population level parameters – $\{(\bar{\alpha}_{j0}, \bar{\alpha}_{j1}, \bar{\alpha}_{j2}, \{\bar{\phi}_{jj'}, \bar{\lambda}_{jj'}, \bar{\mu}_{jj'}\}_{j'=1,2,...,J}), \Sigma\}$ are taken from standard conjugate hyperdistributions, namely –

$$(\bar{\alpha}_{j0},\bar{\alpha}_{j1},\bar{\alpha}_{j2},\{\bar{\phi}_{jj'},\bar{\lambda}_{jj'},\bar{\mu}_{jj'}\}_{j'=1,2,...,J})' \sim N((\bar{\alpha}_{j00},\bar{\alpha}_{j01},\bar{\alpha}_{j02},\{\bar{\phi}_{0jj'},\bar{\lambda}_{0jj'},\bar{\mu}_{0jj'}\}_{j=1,2,...,J})',\Lambda_0);$$
$$\Sigma \sim Wishart(\varphi_0,\Sigma_0)$$

The full conditional posterior distributions of these population level parameters reduce to Normal, Gamma and Multinomial distributions.

Optimal Coupon Targeting Problem

With the increased importance of customized targeting of customers in the retail sector, one of the most important decisions encountered by retailers is to decide on the specific targeting strategies they need to use to reach each customer and subsequently increase the value of the customer over his/ her lifetime. Rather than implementing a point of purchase coupon targeting strategy, which essentially takes into account the basket-mix on that purchase occasion, it is becoming increasingly important for retailers to take into consideration the entire purchase history of the customers as well as form expectations about the future customer responses to their targeting effort.

Here we propose a customized targeting problem where the retailer chooses to target customers with coupons which have unique face values. In this framework, the retailer mails individualized coupons to customers at different time periods based on their knowledge of the customer’s response patterns, customer decides whether to make a brand purchase and also whether to redeem the brand-specific coupon. The retailer’s decision problem has three crucial pillars – which customers to send a coupon, when to send a coupon and finally for which brand the coupon should be mailed. Given the fact that the face value of a coupon send to a customer can be zero, we can effectively consider the retailer’s problem as two-dimensional one where they decide on the face value and the timing of coupon mailing.

Similar to RMA, we also focus on the revenue side of couponing, and assume that the costs associated with coupon redemption and issuing of coupons, even though important, is less of a concern for the retailer since usually retailers look to maximize the returns form a customized couponing. We also assume that customers stockpiling costs are negligible, thus there is no constraint on the timing of coupon redemptions for the customers. If the face value of a coupon is worth a purchase, then the customer will buy the brand offering the coupon. The retailers on their part seek to weigh the benefits of a customized targeting strategy vis-à-vis that of a broad aggregated targeting, primarily in terms of in-store promotional efforts. In order to make a decision regarding the optimal face value of coupons for each customer, the retailer needs to maximize the incremental revenue which they can obtain by providing a coupon compared to a scenario where they charge the regular price from the customers. The ultimate objective is not to acquire customers who would never make a purchase in the product category or to retain customers who would anyways make a purchase. Rather it is important for the retailer to get maximum number of customers to switch brands using the coupons. Retailers aim to shift customers from low margin brands to high margin brands such that their overall returns are maximized.

As discussed earlier, we will approach the coupon targeting problem as a two step optimization framework, one which reformulates the utility function that is maximized and the other one focuses on the learning in a multi-period decision making. First, we need to understand whether there is any additional value to the retailer of taking into consideration the information gain from the targeting, in addition to the maximization of incremental revenue. The important question is – Can learning through a targeting mechanism which takes the form of a natural experimentation actually lead to better targeting? Second, it is important to address the time dimension of the optimization problem. Is a single period optimization sufficient to determine the optimal targeting strategy using customized coupons? Or, is it important to construct a multi-period objective function? Also, if targeting does improve by maximizing the value of customers over multiple periods, is it important to update beliefs based on information gained during the targeting process, thereby taking into considering the expected brand purchase behavior of customers in the future for making targeting decisions in the current period? Such sequential decision process is expected to be more robust than a myopic approach which assumes that the customer is likely to respond in a consistent way over his/ her entire lifetime. Finally, the obvious question is whether altering the retailer's utility function by simultaneously considering incremental revenue and information gain as well as positioning the coupon targeting problem as a sequential decision process further improves the efficiency of the optimal targeting.

The benchmark targeting approaches which we would consider to evaluate the relative performance of our three alternative mechanisms are – A framework where targeting of customers by the retailers is based on a incremental revenue maximization at every period in the decision time horizon. This neither takes into account the value of information, nor does it consider a multi-period problem where expected customer behavior is important. Our second benchmark model considers a multi-period utility function for the retailer for targeting purposes, but only considers expected incremental revenue over a finite time horizon. Thus retailers behave myopically and do not update their beliefs over time.

We would address how each step can be implemented using an algorithm which not only addresses the computational issues arising from the uncertainty over parameters and the retailer's objective function, but also in the evaluation of the expected utilities. We use forward simulation in conjunction with backward induction and a Monte Carlo integration method to significantly reduce the computational burden in the evaluation of the utilities. Also, in any optimal decision making problem which takes into account the expectation of future response behavior in addition to past history encounters the scenario where increasing number of possible decision paths slow down the decision making process. The use of a sufficient statistic in a constrained action space takes care of this concern. Finally since our framework requires update of the retailer's beliefs over time, we need to update the posterior distributions of the parameters using a particle filter where importance weights are proportional to the likelihood.

Assume that the retailer's decision involves determining the optimal face value of coupons for each customer over the decision time horizon $t = T+1,...,\tilde{T}$ once the customer's response model of brand choice has been estimated for periods $t = 1,...,T$. In terms of notations, the decision problem can be written as –

$$\{v_{ijt}^{*}, f_{ijt}^{*}\} = \arg\max E_{\Theta_{ij}}[U(v_{ijt}, f_{ijt}, \Theta_{ij}) \mid y_{ijt}]$$

Here $v_{ijt} = 1$ if the retailer decides to mail a coupon for brand *j* to customer *i* at time *t*; 0 otherwise $(t = T+1,...,\tilde{T})$ and f_{ijt} is the face value of the coupon send. Note that since the face value of the coupon can be zero, thus implying that no coupons have been send, the decision problem reduces to –

(14) $$f_{ijt}^{*} = \arg\max E_{\Theta_{ij}}[U(f_{ijt}, \Theta_{ij}) \mid y_{ijt}]$$

Also note that $\Theta_{ij} = \{\alpha_{ij0}, \alpha_{ij1}, \alpha_{ij2}, \{\phi_{ijj'}, \lambda_{ijj'}, \mu_{ijj'}\}_{j' \neq j = 1,2,...,J}\}$ are the customer level posterior distribution of parameters obtained by estimating the customer response model of brand purchases, f_{ijt}^{*} is the optimal coupon face value for customer *i* at time *t* for brand *j* and $U(f_{ijt})$ is the utility function of the retailer for the Bayesian decision making process. Note that the decision time horizon is $t = T+1,...,\tilde{T}$

When we consider the benchmark models, the utility function for the first benchmark model of the retailer is based on the incremental sales for the retailer for using a coupon with temporary price reduction. This is given by –

(15) $$\Pi^{1}_{ijt} = Pr\left(y_{ijt} = 1 \mid \Theta_{ij}, p_{ijt} - f_{ijt}, \tilde{X}_{ijt}\right) - Pr\left(y_{ijt} = 1 \mid \Theta_{ij}, p_{ijt}, \tilde{X}_{ijt}\right)$$

for $t = T+1,...,\tilde{T}$

Here $\tilde{X}_{ijt}$ are all the time-varying factors which affect the hazard of brand purchase which depends on the duration from previous purchase occasion, apart from the price of the own brand, i.e. cross brand prices, dummy variables indicating targeting by the retailer with coupons for each brand as well as in-store promotion dummies for all brands within the product category and $Pr\left(y_{ijt}=1\,|\,.\right)$ denotes the purchase probability for brand *j* at time *t*.

Note if we assume that g_{ijt} is the margin to the retailer for each unit of brand *j* sold at time *t* to customer *i* with f_{ijt} being the face value of the coupon being offered at time t to customer *i* for brand *j*, the net incremental revenue to the retailer resulting from coupon usage can be written as –

$$\text{(16)} \qquad \Pi^{1}_{ijt} = Pr\left(y_{ijt}=1\,|\,\Theta_{ij}, p_{ijt}-f_{ijt}, \tilde{X}_{ijt}\right)\left(g_{ijt}-f_{ijt}\right) - Pr\left(y_{ijt}=1\,|\,\Theta_{ij}, p_{ijt}, \tilde{X}_{ijt}\right) g_{ijt}$$

this would be used as the first benchmark which we would consider when comparing our optimal sequential coupon targeting algorithm.

The second benchmark model for comparison purposes is given by –

$$\text{(17)} \qquad \Pi^{2}_{ijt} = \sum_{t=T+1}^{\tilde{T}} \kappa^{t} \Pi^{1}_{ijt} \;\text{ for }\; t = T+1, \ldots, \tilde{T}$$

Here κ is a discount factor with $0<\kappa<1$.

Now in our optimal coupon targeting framework, we not only consider the increase in the incremental purchase probability in the product category, but also how we can gain information as part of the targeting process. This idea is in essence similar to a natural experimentation method. Let us consider the following example. Suppose a retailer has a sample of 1000 customers to whom he/ she can potentially send customized coupons for certain brands within a product category. In period 1, after sending coupons to all 1000 customers, the retailer observes a 35% redemption rate which only account for a fraction of the total purchase occasions. The rest might be customers who would have never purchased in that product category anyways or certain customers who would make a purchase irrespective of the coupon send, and so they did not redeem the coupon even when making a purchase. With some information gained in period 1 regarding the customers' purchase behavior, in period 2, the retailer might actually send out coupons to 800 customers and improve the redemption rate to 45%. Again, based on the information gained in period 2, 600 customers might be targeted in period 3 with a redemption rate of 55% and so on. So effectively the retailer is able to narrow down on the group of customers who need to be targeted using the coupons, thus making the marketing strategy more efficient. Since the customer does not necessarily redeem a coupon whenever he/ she makes a purchase, if a retailer can increase the probability of redeeming a coupon on any purchase occasion, it would serve dual purposes – first, it would automatically increase the purchase of the focal brand and second, it would increase the possibility of a brand switching which might lead to a overall increase in the sale of the high margin brand relative to that of a lower margin brand, thus increasing category profits for the retailer. Overall, this might require the retailer to sacrifice some incremental revenue in the initial periods, but in the long run revenue is maximized when the information gain is also optimal.

Value of Information:

To capture the information gain at every stage of the decision making process, we use the Shannon Information Criteria (SIC) (Shannon, 1948), which measures the updating of information regarding the customer's response patterns with the log-ratio of posterior distributions of the parameter vector in decision time period *t* over time period *(t-1)*. This was we can reduce the uncertainty in the parameters by capturing the information at each stage of the decision making process, thus reducing uncertainty in the targeting. The SIC can then be written as the following –

$$S_{ijt} = \log\left[\frac{p\left(\Theta_{ij} \mid y_{ijt}, f_{ijt}\right)}{p\left(\Theta_{ij} \mid y_{i,j,t-1}, f_{i,j,t-1}\right)}\right] \text{ for } t = T+1,\ldots,\tilde{T} \tag{18}$$

Then following Verdinelli and Kadane (1992), we specify the utility function of the retailer as follows –

$$U\left(f_{ijt}, \Pi_{ijt}, \Theta_{ij}, S_{ijt}, \chi_{it}\right) = E_{\Theta_{ij}}\left[\Pi_{ijt} + \chi_{ijt} S_{ijt}\right] \quad \text{for } t = T+1,\ldots,\tilde{T} \tag{19}$$

Here χ_{ijt} measures the contribution of information gain relative to the incremental increase in purchase probability in the retailer's utility function due to the purchase of brand *j* at time *t* by customer *i*. As explained earlier, this objective function is the first step in the targeting decision of the retailer where augment the utility where we add the value of information to the increment in revenue which results from coupon usage. Augmenting the retailer's utility function in this way requires us to update the posterior parameter distributions at every stage of the decision making process which is addressed by using a particle filter with importance weights proportional to the data likelihood. This is first model which needs to be compared against the benchmark models.

Sequential Bayesian Decision Theoretic Algorithm:

Now we add another level of complexity to the decision making process of the retail marketer by considering a sequential decision making process to determine the optimal coupon face value. Here the retailer does not initially consider gain in information in addition to the increment in the probability of purchase, but rather considers the utility function defined in equation (16) and sequentially updates the beliefs over multiple periods in order to arrive at the optimal decision. The algorithm explained here was originally proposed by Muller et al. (2006). They considered a simulation-based method for exploration and maximization of expected utility in sequential decision problems. The problem requires backward induction with analytically intractable expected utility integrals at each stage. Muller et al. (2006) proposed to use forward simulation to approximate the integral expressions, and a reduction of the allowable action space to avoid problems related to an increasing number of possible trajectories in the backward induction. The artificially reduced action space allows strategies to depend on the full history of earlier observations and decisions only indirectly through a low dimensional summary statistic. The proposed rule provides a finite-dimensional approximation to the unrestricted infinite-dimensional optimal decision rule. The steps involved in the optimization algorithm are as follows –

1) The customer response model for purchase decisions at the brand level are estimated using Bayesian MCMC methods for periods $t = 1,...,T$ and the posterior distributions of the parameters are evaluated.

2) The decision making problem involves the determination of optimal coupon face value in periods $t = T+1,...,\tilde{T}$. To get $\Pi_{ij,T+1}$, we first draw $\Theta_{i,j,T+1} \sim p(\Theta_{ij})$ where Θ_{ij} is the posterior distribution of the parameter vector for the customer brand choice model. Then the posterior predictive distribution of Π_{ij} ($p(\Pi_{ijt} \mid \Theta_{ij})$) is used to simulate pseudo-observations $\Pi_{i,j,T+1} \sim p(\Pi_{i,j,T+1} \mid \Theta_{i,j,T+1})$.

3) However, in order to obtain the posterior full conditional distribution of parameters in period (T+1), we need to use a particle filter which updates the posterior distribution of the parameter vector using an importance weighting scheme where the importance weights are proportional to the likelihood of the data, i.e. the incremental probability resulting from the use of coupons by the customers. This idea was first proposed by Ridgeway and Madigan (2002). Accordingly we get –

(20)
$$p(\hat{\Theta}_{i,j,T+1} \mid \Pi_{i,j,T+1}) \approx \frac{\sum_{m=1}^{M} w_m \delta(\Theta_{ij} - \Theta_{ij}{}^{m})}{\sum_{m=1}^{M} w_m}$$

where *m=1,…,M* are the no. of particles; δ is the Dirac delta measure and w_m is the importance sampling weight with $w_m \propto L(\Pi^m \mid \Theta^m)$

4) Then we draw $\Theta_{i,j,T+2}$ from the posterior distribution using the inverse-CDF method from a Uniform distribution.

5) Subsequently we simulate $\Pi_{i,j,T+2} \sim p(\Pi_{i,j,T+2} \mid \Theta_{i,j,T+2})$ and proceed similarly as before. Note that so far we have not taken into account the value of information in the utility function specification for the retailer. Without gaining market intelligence in terms of the information update, the decision of the retailer is strictly contingent on the incremental probability of purchase in the product category. This might lead to higher gains in the initial periods of the decision time horizon, but over the lifetime of the customers, the retailer might have lower overall returns from their targeting strategies since there is no learning.

6) Then we add the Shannon Information Criteria to the retailer's utility function which now takes the functional form defined in equation (19). The next step in the retailer's decision making process involves the evaluation of the SIC given by S_{ijt}. Since we have already used the particle filtering approach to update the posterior distribution of the parameters in periods $\hat{\Theta}_t \left(t = T+1,...,\tilde{T}\right)$, S_{ijt} can be evaluated by taking the log-ratio of the posterior distributions in *t* and *(t-1)*.

7) To draw the relative weight of information gain vis-à-vis the increment in purchase probability, we run *q=1,…,Q* such simulated experiments within a 2-dimensional grid for χ_{ijt}. This enables us to define the weight in a restricted domain within certain bounds of the incremental probability and information gain.

8) Then for each simulated experiment q, we record K_{ijt}^{q} $(t = T+1,...,\tilde{T})$ where K_{ijt}^{q} is a sufficient statistic defined as – $K_{ijt} = K_{ijt}(U_{ij}^{t-1} \mid f_{ij}^{t-1}) = \{m_{ijt}, s_{ijt}\}$. Here

$(U_{ij}^{t-1} = U_{ij1},....,U_{i,j,t-1},\ f_{ij}^{t-1} = f_{ij1},....,f_{i,j,t-1}); m_{ijt} = E(U_{ij}^{t-1})$ and $s_{ijt} = E(U_{ij}^{t-1} - m_{ijt})^2$

which are the mean and variance of the utility function for the retailer over the decision time horizon prior to period *t*. As explained earlier, it important to use a sufficient statistic since with the increasing number of possible trajectories in a sequential decision problem, the optimization algorithm using a backward induction method leads to an explosion in the number of decision paths. So to summarize the information contained in all the possible decision paths, Muller et al. (2006) proposed the use of a sufficient statistic which captures the entire history of earlier actions in terms of a lower dimensional measure.

9) As the final step of the optimization algorithm, we again follow Muller et al. (2006). To handle the analytically intractable expected utility integrals they propose the use of forward simulation for the evaluation of the expected utilities after integrating out the parameter uncertainty, in conjunction with backward induction and use a grid based method to determine the optimal decision paths.

The method is showed in Figure 4. On the horizontal axis we take the decision time horizon while on the vertical axis we have a one-dimensional sufficient statistic (for purposes of easier explanation). Now over the entire decision time horizon we observe numerous decision trajectories. Let us focus on sets of paths in dark bold lines. In the backward induction method, as we start from the last time period (period 10 in this case), we see that all of the dark bold lines pass through a particular grid cell (say l_1). The grid based optimization works the following way – We determine the optimal decision path in period 10 for cell l_1 by maximizing over all the possible paths in grid cell l_1. This can be done for all cells in period 10. Once we have determined the optimal decisions made in period 10 for all grid cells, we move back one period to period 9. Then for any grid cell l_2 for that period we consider the trajectories which pass through that cell. Subsequently, given the evaluated value of the optimal decisions made in period 10 for the cells through which these decision paths have passed, we determine the optimal decision for grid cell l_2 in period 9. This process continues backward until we arrive at the first period of the decision time horizon.

First we construct a three dimensional grid (t, K_{ijt}) for $t = T+1,...,\tilde{T}$ where K_{ijt} has two dimensions as explained above. The optimization starts from period $\tilde{T}$ by evaluating $\hat{U}_{ij\tilde{T}}(f_{ij\tilde{T}}, K_{ij\tilde{T}} = \{m_{ij\tilde{T}}, s_{ij\tilde{T}}\}) = \frac{1}{M_{ij\tilde{T}_l}} \sum_{q \in A_{ij\tilde{T}_l}} U(f_{ij\tilde{T}}, \Pi_{ijq}^{\tilde{T}}, S_{ijq}^{\tilde{T}}, \Theta_{ijq})$ for all $\tilde{T}$. Note that $A_{ij\tilde{T}_l}$ is the subset of $q \in \{1, 2, ..., Q\}$ corresponding to all trajectories or decision paths which terminate in cell l of the three dimensional grid and $M_{ij\tilde{T}_l} = |A_{ij\tilde{T}_l}|$ is the number of indices in $A_{ij\tilde{T}_l}$. Then from $\hat{U}_{ij\tilde{T}}(f_{ij\tilde{T}}, K_{ij\tilde{T}})$ we find $f^*_{ij\tilde{T}}(K_{ij\tilde{T}})$ as the optimal for each grid cell *l*.

Hence we can write - $\hat{U}^*_{ij\tilde{T}}(K_{ij\tilde{T}}) = \hat{U}_{ij\tilde{T}}(f^*_{ij\tilde{T}}, K_{ij\tilde{T}})$. Proceeding in the same way for

$t = \tilde{T} - 1, ..., T + 1$, we approximate expected utility for each period *t* as -

$\hat{U}_{ijt}(f_{ijt}, K_{ijt}) = \frac{1}{M_{ijt_l}} \sum_{q \in A_{ijt_l}} \hat{U}^*_{i,j,t+1}(f_{i,j,t+1}, \Pi_{i,j,t+1,q}, S_{i,j,t+1,q})$ where

$\hat{U}^*_{i,j,t+1}(f_{i,j,t+1}, \Pi_{i,j,t+1,q}, S_{i,j,t+1,q})$ is the optimal decision for period (t+1) and A_{ijt_l} is a subset of size M_{ijt_l} of all trajectories that pass through cell *l*. Hence the optimal decision at any time period *t* is given by $f_{ijt}^*(K_{ijt})$ and the optimal utility is written as –

$\hat{U}^*_{ijt}(K_{ijt}) = \hat{U}_{ijt}(f^*_{ijt}, K_{ijt})$.

Concluding Remarks

This research proposes a theoretical optimal Bayesian Decision Algorithm for decision making in the context of retail sector. The decision involves the individual customer level targeting using coupons where the objective of the retailer is threefold – first he/ she need to decide which customers to send a coupon and if so what should be the face value of the coupon; the second decision involves the timing of the coupon mailing and finally the retailer chooses which brand’s coupon to send to a particular customer. We first construct the customers’ brand choice problem as a competing risk discrete hazard model, which takes into account not only the prices and in-store promotional strategies of retailers, but also how duration from past purchases, historical customer targeting using coupons and the coupon face values affect the customer's brand level purchase behavior. In the coupon targeting decision process of the retailer, we propose an algorithm which addresses two important issues which has so far not been considered in the marketing literature – First, we explore how we can potentially enhance the decision making process by not only including the incremental revenue resulting from coupon redemption during brand purchase in the retailer's utility function, but also considering the role of information in the optimal outcomes; Second, we construct a multi-period targeting problem that takes into account a sequential update of beliefs over the customer's lifetime. This incorporates the potential evolution of customer's response behavior with respect to their brand purchase behavior resulting from repeated interactions with the retailer. This work provides a valuable framework to retailers in understanding the brand purchase behavior of their customers, and more specifically how category purchases are affected by brand level coupon redemptions in conjunction with pricing, coupon targeting and in-store promotions, how retailers can use sophisticated decision theoretic approaches to determine optimal coupon targeting strategies and finally whether there is value of information as well as relevance of sequential learning in the decision making process.

If we compare the benchmark optimal decision frameworks with the three proposed alternatives (namely – modification of the retailer's utility function to include gain in information in addition to incremental revenue of a couponing scheme, but looking at a one-period ahead targeting decision (Alternative A); sequential decision making where retailer's update their beliefs over time, but only consider the incremental revenue (Alternative B) and finally a multi-period sequential targeting where optimal decision take into account the trade-off between revenue increments and information gain (Alternative C)), we expect the following. First, Alternative C is surely going to improve the efficiency of the optimal targeting with respect to the determination of coupon face-values compared to either of the benchmark models as well as Alternatives A and B since it not only considers an utility function that explicitly models the information gain, but also allows for the updating of beliefs over time in the sequential set-up. Second, Alternative B is also likely to improve on the efficiency of the benchmark model where we consider a multi-period optimal targeting problem, albeit without any update of beliefs on the retailer's part. Obviously, it is also expected to perform better than the single period decision making problem. This is because learning through updating of beliefs would help the retailers adapt to the evolving response patterns of the customers while in the baseline approach, retailers are essentially myopic and assume that customers are going to be consistent in their brand purchase behavior. The performance of Alternative A compared to the baseline approaches is expected to be mixed. When compared to the baseline model which just considers single-period profit maximization for optimal targeting we expect Alternative A to perform better, for the baseline model which maximizes the expected revenue over a finite time horizon, the result can go either way. This second baseline model scores by constructing a multi-period decision problem which takes into account the customer's evolution in future behavior. Alternative A on the other hand provides a more robust approach when making decisions one period at a time by augmenting the utility function with the gain in information.

One of the potential shortfalls of this research are the problems related the computational costs in the optimal decision making algorithm. Even though the customer response model for brand choices is relatively easier to implement using Bayesian estimation methodologies, the sequential decision process of the retailer, specially with the augmented utility function that includes both incremental returns from coupon usage and information gain, is resource intensive. If the decision time horizon is large, the computational problems are intensified even further. As we have discussed earlier, the backward induction algorithm to solve the sequential decisions not only leads to exponentially increasing number of possible decision paths, but also requires the evaluation of analytically intractable integrals and optimizations over multiple dimensions. The solutions to address these problems proposed in this research do reduce the burden significantly, but it still remains a concern. However, the benefits of the proposed method surely outweigh the costs and we believe that with the increase in the computational power of modern day computers, these issues can be handled relatively easily.

CHAPTER FOUR: DISCUSSION ON IMPLICATIONS AND FUTURE RESEARCH DIRECTIONS

Background

The overarching focus of this dissertation was to explore the opportunities available to marketers to target customers using direct marketing tools in two sectors of the economy which has traditionally not seen much research in this area. Both the non-profit sector and retail sector are important pillars of the economy, though researchers have not focused on these markets when building models of customer response to direct marketing initiatives and subsequently ignored how optimal targeting mechanisms can be formulated to maximize customer value in these sectors. In this dissertation, I try to focus exclusively on these two sectors and build customer response models addressing two closely related yet very different direct marketing contexts. The objective was to provide marketers with empirical modeling tools which would can not only equip them with the technical knowhow to understand customer response patterns, but also can be translated into actual implementation of optimal direct marketing strategies.

Theoretical and Managerial Implications of First Essay

In the case of the non-profit sector , which is the application domain of my first essay, the customer response model tries to understand the donation behavior of customers (donors) in terms of the timing and amount of contributions and provides insights into the role played by past donation patterns (duration dependence and volume dependence) and solicitation strategies on current contributions. This approach also explores the underlying heterogeneity in donation behavior by explaining the differences in donations with individual and community level donor characteristics. The framework allows us to create donor portfolios using the frequency and amount of contributions and then characterize these portfolios with donor characteristics. Such characterization helps the identification of donor groups simply from their profiles, rather than directly observing their true donation behavior or having to collect information regarding their intensions to donate.

Though we use a sophisticated multi-episode discrete proportional hazard model to formulate the timing of donations in conjunction with a log-Normal regression to formulate the donation amounts and estimate these hierarchical models using Bayesian MCMC methods, from an implementation point of view, they are simple and can be easily interpreted by marketing managers from the non-profit sector. In the empirical application we use a donation dataset which not only capture the essence of the model, but also provides very interesting take-aways regarding donor response behavior. We see that donation behavior is not only strongly duration dependent, but also past donation amounts as well as the pattern of past solicitations strongly influence current donations. Additionally, there are discernable differences across different donation programs. More specifically, across two different donation programs we showed that the donor characteristics which matter in explaining the heterogeneity in donor behavior changes significantly. This provides us with the intuition that non-profits can potentially identify donor segments using their characteristics since the portfolios for each donation program are unique.

The fact that we can create profiles of donors by characterizing donor portfolios with their characteristics provides the direct marketer with a heuristic approach to target donors when their true behavior is unknown. In most cases when non-profit organizations seek to attract new donors who can potentially contribute to the causes sponsored by the non-profits, they lack information regarding the donors' true potential in terms of donation patterns. However, if the managers can see the donor characteristics, they can extrapolate the expected donation behavior by matching the donor characteristics with the donor portfolios. Even though this approach is not full proof and we see that it might lack some accuracy under limitations of data, theoretically it is appealing since it is very easily implementable.

Limitations of First Essay and Avenues for Future Research

We must also point out here that there are limitations in our first essay which needs to be acknowledged. First and foremost, the dataset we are using for the empirical application of the donation response model and the subsequent heuristics for donor targeting is relatively sparse. So, we only observe actual donation incidences for a few periods for each donor and this surely affects the estimation results.

Also, in order to explain the observed heterogeneity in donation behavior, we have collected data from external sources on the individual and community level donor characteristics. Given the lack of donor-specific data, we had to rely on publicly available sources which are often aggregated to begin with. Thus even though ideally we would have preferred to have donor characteristics available at the individual level, and we could summarize this information by aggregating the information ourselves, we cannot capture a significant portion of the heterogeneity. So even though the idea regarding the matching of donor portfolios with their corresponding characteristics is theoretically appealing, the performance of the prediction exercise does not do well as per our expectations.

To address the above two limitations of our research, future work can apply our modeling framework to a better dataset which not only has more actual donation incidences, but also provides information regarding the donors' characteristics at a more disaggregated level. This would substantially improve the model's performance which is important since the non-profit sector has surely got potential for direct marketing efforts to succeed. Additionally, our modeling approach in the first essay makes a few important assumptions. First of all, we assume that donation to one cause does not prevent the donor from contributing to another cause, thus effectively ruling out competing risks given that there are no budget constraints. In the second essay, which is positioned in the context of retail brand purchase problem, we explicitly model competing risks. Second, we assume that the costs for the direct marketer in the non-profit sector are negligible. So they can target donors with the heuristic targeting approach without bothering about the costs associated with sending solicitation mails. Third, even though we consider community level donor characteristics to explain the heterogeneity in donation behavior, the true essence of the community structure might not be captured in their demographics, financial well being etc. Rather it is possible that culture, societal norms, moral values and even philanthropic actions undertaken by fellow community members might play a more role in explaining the differences in donation behavior. Also the donation behavior of other members of the community might also be critical. Most of these limitations can be addressed easily handled in future research.

Theoretical and Managerial Implications of Second Essay

The second essay shifts our attention to another interesting sector of the economy, namely the retail sector, where direct marketing and more specifically individual customer-specific direct marketing strategies have been virtually overlooked. Building up on one of the limitations pointed out for the first essay, namely the customer response model using competing risk hazard framework, we model the purchase behavior for each brand within a product category in the retail sector. Redemption of coupon for any brand, pricing, in-store promotions, the targeting by retailers using coupons as well as coupon face values impact the brand choice. For implementation in an empirical context, the competing risk discrete hazard model can be simplified under the assumption that transitions between one state to another can only happen at the boundaries of the interval between episodes. This way a competing risk model of brand choice can be estimated in terms of the individual discrete hazard models of brand choice.

With the understanding of the customer response model in terms of brand purchase, the second essay adopts a structural algorithm for direct marketing in the retail sector. The decision process of the retailer requires them to decide which customers to target, with which brand's coupon, at what face value and also when to target. Such customized targeting in the retail sector is relatively unique. In direct marketing, specially in the context of catalog marketing, various decision theoretic approaches have been proposed for optimal targeting. However, unlike most research currently existing in the marketing literature, we make two crucial amendments to the decision problem. First, we propose that in making optimal decisions regarding coupon targeting, retailers need to take into account the information gain from each targeting event in addition to the incremental revenue they are expected to earn due to the usage of coupons. This is important because with the additional term in the retailer's objective function, the retailer might need to sacrifice some revenues in the initial periods of the decision time horizon in order to increase the flow of revenue in the future over the customer's lifetime. Second, we also believe that using a multi-period decision problem does not suffice to increase the efficiency of optimal targeting with coupons in the retail sector. Unless the retailers are myopic and assume the customers are not likely to evolve in their response behavior over their lifetime, a lifetime value calculation over a finite time horizon as a discounted present value of future revenue streams might be flawed. On the other hand, if we assume that the retailer can learn from the repeated targeting and accordingly update their beliefs sequentially in each period, it would allow the retailers to adapt as the customer evolves.

We propose an algorithm using a Bayesian decision theoretic approach which takes into account the above mentioned managerially important insights. However, this leads to computational complexities in the optimal coupon targeting problem. Given that our approach is sequential, it leads to increasing number of possible decisions paths. Also, the functional form of the retailer's utility function requires us to update the posterior parameter space at every stage of the decision making process and evaluate intractable expected utility integrals. Our algorithm provides solutions to each of these problems – namely a sufficient statistic over a constrained action space takes care of the problem of exponentially increasing decision oaths; a particle filter continuously updates the parameter space at each stage of the decision time horizon and the use of a Monte Carlo forward simulation in conjunction with backward induction leads to the numerical evaluation of the expected utility integrals.

Limitations of Second Essay and Avenues for Future Research

Unfortunately the mathematically demanding optimal targeting algorithm in the second essay automatically restricts the scope for empirical application. Even after providing solutions to the pressing issues in the algorithm, it still remains computation intensive and the higher the dimension of the decision time horizon, the feasibility of the application becomes debatable. Further, since customized targeting to determine optimal coupon face values in the retail sector is still very new as a possible direct marketing strategy, there are no suitable datasets currently available to empirically validate the performance of the Bayesian decision theoretic approach to optimal targeting against other benchmark approaches. The coordination between the retailers and the manufacturers in order to provide the customized coupons might also turn out to be a challenge when we consider the implementation of the algorithm in a field experiment. Possible future research can not only test this validity using simulated datasets, but also should look to collaborate with retail stores and devise the optimal targeting strategies for field testing and empirical validation. We are confident that such structural approaches for targeting customer directly in the retail sector would provide an excellent opportunity to the retailers who have traditionally tried to reach customers either using generic coupons through mail or in-store point of purchase coupons. The overall efficiency of the targeting would be improved significantly since over time wastage can be avoided when customers are not expected to respond favorably to the targeting efforts of the retailer.

TABLES

Table 1. Classification of Donors

	Amount of Contributions (in $)		
Frequency of Contributions		*Low*	*High*
	Low	**Donor Segment A**	Donor Segment B
	High	Donor Segment C	**Donor Segment D**

Table 2. Descriptive Statistics Related to Donation Behavior

Estimation Period Characteristics	***Donation Program 1 (1130 donors)***		***Donation Program 2 (281 donors)***	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*
Number of Donations	**3.99**	2.53	**2.91**	1.15
Number of Solicitations	**5.30**	1.71	**1.83**	1.09
Duration between Contributions (in weeks)	21.95	15.68	39.36	23.81
Frequency of Contribution	**0.063**	0.03	**0.036**	0.02
Duration between Solicitation and Contribution (in weeks)	2.50	2.83	2.85	3.31
Amount of Donation per Donation Incidence (in $)	**8.97**	7.61	**13.58**	11.45

Table 3. Mean (Standard Deviation) for Individual and Community Donor Characteristics

Donor Characteristics	***Donation Program 1***	***Donation Program 2***
Individual Level		
Gender (Fraction of Male)	**0.57 (0.49)**	**0.62 (0.49)**
No. of Lifetime Solicitations	**20.94 (11.27)**	**26.11 (10.55)**
ZIP Code Level		
Median Age (in years)	45.75 (5.70)	46.07 (5.76)
Percentage of White	73.25 (20.95)	72.58 (21.62)
Median Years of Schooling	13.26 (1.22)	13.25 (1.28)
Persons Per Household	**2.76 (0.38)**	**2.57 (0.38)**
Percentage of Households which are Families	72.77 (10.56)	72.37 (10.12)
Estimated Median Income (in 1000 $)	53.74 (21.17)	54.20 (23.41)
Median Home Value (in 1000 $)	**76.52 (45.38)**	**78.51 (49.91)**
Total No. of Tradelines	13.37 (2.14)	13.27 (2.19)
No. of Protestant Churches per Capita	0.0007 (0.0007)	0.0008 (0.001)
No. of Non-Protestant Churches per Capita	0.0003 (0.0003)	0.0003 (0.0004)
Average Rating of Schools	2.87 (0.59)	2.85 (0.61)
County Level		
Percentage of Republican Votes	61.10 (10.95)	61.51 (11.42)
Total Crime per Capita	**0.0182 (0.06)**	**0.0155 (0.05)**

Table 4a. Donation Behavior for Program 1 Donors for each Segment

		Amount of Contributions (in $)	
		Low	***High***
Frequency of Contributions	***Low***	*Donor Segment A*	*Donor Segment B*
		No. of Donors – 335 Total Number of Contributions – 2.64 **Frequency of Contributions – 0.04** Duration between Contributions (in weeks) – 27.77 **Amount of Contributions (in $) – 5.15**	**No. of Donors – 256** Total Number of Contributions – 2.39 **Frequency of Contributions – 0.03** Duration between Contributions (in weeks) – 35.80 **Amount of Contributions (in $) – 16.34**
	High	*Donor Segment C*	*Donor Segment D*
		No. of Donors – 374 Total Number of Contributions – 5.82 **Frequency of Contributions – 0.09** Duration between Contributions (in weeks) – 11.40 **Amount of Contributions (in $) – 4.92**	**No. of Donors – 165** Total Number of Contributions – 5.03 **Frequency of Contributions – 0.08** Duration between Contributions (in weeks) – 12.58 **Amount of Contributions (in $) – 14.56**

Table 4b. Donation Behavior for Program 2 Donors for each Segment

		Amount of Contributions (in $)	
		Low	***High***
Frequency of Contributions	***Low***	*Donor Segment A*	*Donor Segment B*
		No. of Donors – 99 Total Number of Contributions – 1.23 **Frequency of Contributions – 0.02** Duration between Contributions (in weeks) – 50.55 **Amount of Contributions (in $) – 7.34**	**No. of Donors – 67** Total Number of Contributions – 1.30 **Frequency of Contributions – 0.02** Duration between Contributions (in weeks) – 57.54 **Amount of Contributions (in $) – 23.34**
	High	*Donor Segment C*	*Donor Segment D*
		No. of Donors – 78 Total Number of Contributions – 2.49 **Frequency of Contributions – 0.06** Duration between Contributions (in weeks) – 19.07 **Amount of Contributions (in $) – 7.66**	No. of Donors – 37 Total Number of Contributions – 2.97 **Frequency of Contributions – 0.06** Duration between Contributions (in weeks) – 19.25 **Amount of Contributions (in $) –25.09**

Table 5. Identifying Donors – Features of Each Donor Segment for Program 1

Total Donors = 1130	***Amount of Contributions (in $)***		
		Low	***High***
Frequency of Contributions	***Low***	*Donor Segment A* *No. of Donors = 335*	*Donor Segment B* *No. of Donors = 256*
		Gender (fraction of male) – 0.54 **Lifetime Solicitations – 17.48** **Age (in years) – 46.08** White (%) – 72.15 **Schooling (in years) – 13.19** Household Size – 2.76 Family (%) – 72.97 **Income (in 1000 $) – 52.55** **Home Value (in 1000 $) – 74.79** **Tradelines – 13.20** **Per Capita Protestant Churches – 0.0007** Per Capita Non-Protestant Churches – 0.0003 **Education Rating – 2.84** **Republican Votes (%) – 61.64** Per Capita Crime Rate Crime – 0.02	**Gender (fraction of male) – 0.61** **Lifetime Solicitations – 22.08** **Age (in years) – 44.71** White (%) – 74.39 **Schooling (in years) – 13.46** Household Size – 2.73 Family (%) – 71.95 **Income (in 1000 $) – 57.04** **Home Value (in 1000 $) – 83.40** **Tradelines – 13.70** **Per Capita Protestant Churches – 0.0006** Per Capita Non-Protestant Churches – 0.0003 **Education Rating – 2.95** **Republican Votes (%) – 59.70** Per Capita Crime Rate – 0.02
	High	*Donor Segment C* *No. of Donors = 374*	*Donor Segment D* *No. of Donors = 165*
		Gender (fraction of male) – 0.54 **Lifetime Solicitations – 21.90** **Age (in years) – 46.10** White (%) – 73.72 **Schooling (in years) – 13.19** Household Size – 2.75 Family (%) – 72.52 **Income (in 1000 $) – 51.81** **Home Value (in 1000 $) – 72.10** **Tradelines – 13.24** **Per Capita Protestant Churches – 0.0008** Per Capita Non-Protestant Churches – 0.0003 **Education Rating – 2.84** **Republican Votes (%) – 62.18** Crime – 0.02	**Gender (fraction of male) – 0.64** **Lifetime Solicitations – 24.04** **Age (in years) – 45.91** White (%) – 72.65 **Schooling (in years) – 13.29** Household Size – 2.82 Family (%) – 74.18 **Income (in 1000 $)– 55.43** **Home Value (in 1000 $) – 79.42** **Tradelines – 13.50** **Per Capita Protestant Churches – 0.0006** Per Capita Non-Protestant Churches – 0.0003 **Education Rating – 2.86** **Republican Votes (%) – 59.72** Crime – 0.02

Table 6. Posterior Median Parameter Estimates and Significant Effects for Duration Dependence of Donation Behavior **(* denotes significance measured in terms of the posterior probability of observing an effect, given the data and model)**

Time –Varying Factors	***Donation Program 1***		***Donation Program 2***	
	Incidence	***Amount***	***Incidence***	***Amount***
Intercept	**-4.416***	**1.578***	**-6.037***	**2.341***
Duration from Solicitation to Contribution (SolConDur)	**-0.340***	-0.006	**-0.336***	-0.043
Ln(SolConDur)	-0.135	0.069	**-0.526***	0.230
Duration from Previous Contribution (ConConDu)r	**0.029***	**0.033***	**0.108***	**0.052***
Ln(ConConDur)	**0.415***	**-0.409***	**0.481***	**-0.850***
Discounted Value of Previous Contribution Amount (DiscDonAmnt)	**-4.861***	**-1.444***	**-4.925***	**-1.354***

Table 7a. Posterior Median Estimates of Variance-Covariance Matrix (Donation Incidence)

Incidence Model						
Time-Varying Factors	***Donation Program 1***					
	Intercept	*SolConDur*	*Ln(SolConDur)*	*ConConDur*	*Ln(ConConDur)*	*DiscDonAmnt*
Intercept	0.353					
SolConDur	-0.067	0.181				
Ln(SolConDur)	0.305	-0.459	1.637			
ConConDur	0.019	-0.008	0.028	0.026		
Ln(ConConDur)	-0.202	-0.030	0.019	-0.056	0.719	
DiscDonAmnt	0.786	-0.485	1.690	0.066	-0.318	3.400
Time-Varying Factors	***Donation Program 2***					
	Intercept	*SolConDur*	*Ln(SolConDur)*	*ConConDur*	*Ln(ConConDur)*	*DiscDonAmnt*
Intercept	0.812					
SolConDur	0.063	0.163				
Ln(SolConDur)	0.183	-0.070	0.579			
ConConDur	0.024	0.000	0.013	0.071		
Ln(ConConDur)	-0.601	-0.036	-0.168	-0.075	1.155	
DiscDonAmnt	0.942	0.061	0.235	0.076	-1.036	2.167

Table 7b. Posterior Median Estimates of Variance-Covariance Matrix (Donation Amount)

Amount Model						
Time-Varying Factors	***Donation Program 1***					
	Intercept	*SolConDur*	*Ln(SolConDur)*	*ConConDur*	*Ln(ConConDur)*	*DiscDonAmnt*
Intercept	0.237					
SolConDur	0.005	0.046				
Ln(SolConDur)	0.027	-0.048	0.188			
ConConDur	-0.001	-0.001	0.001	0.015		
Ln(ConConDur)	-0.027	0.007	-0.013	-0.017	0.216	
DiscDonAmnt	0.077	-0.005	0.014	-0.003	0.059	0.454
Time-Varying Factors	***Donation Program 2***					
	Intercept	*SolConDur*	*Ln(SolConDur)*	*ConConDur*	*Ln(ConConDur)*	*DiscDonAmnt*
Intercept	0.317					
SolConDur	0.040	0.143				
Ln(SolConDur)	0.005	-0.092	0.433			
ConConDur	-0.004	-0.001	-0.001	0.053		
Ln(ConConDur)	-0.037	0.007	0.003	-0.025	0.419	
DiscDonAmnt	0.038	-0.006	-0.002	0.001	0.029	0.474

Table 8. Posterior Median Parameter Estimates and Significant Effects for Observed Heterogeneity in Donation Behavior **(* denotes significance measured in terms of the posterior probability if observing an effect, given the data and model)**

Donor Characteristics	*Time –Varying Factors*	*Donation Program 1*		*Donation Program 2*	
		Incidence	*Amount*	*Incidence*	*Amount*
Individual Level					
Gender (Fraction of Male)	*Intercept*	-0.082	-0.004	**0.703***	0.033
	SolConDur	-0.082	0.009	**-0.218***	0.189
	Ln(SolConDur)	0.264	0.019	**0.796***	-0.711
	ConConDur	**0.034***	0.016	-0.052	-0.030
	Ln(ConConDur)	**-0.309***	**-0.222***	0.779	0.091
	DiscDonAmnt	-0.294	**-0.629***	**3.818***	-0.340
No. of Lifetime Solicitations	*Intercept*	**0.014***	**0.027***	-0.001	0.004
	SolConDur	**0.009***	0.002	0.009	-0.006
	Ln(SolConDur)	**-0.020***	-0.002	-0.027	0.032
	ConConDur	**0.001***	**-0.002***	0.001	**-0.005***
	Ln(ConConDur)	**-0.042***	**0.026***	**-0.067***	**0.062***
	DiscDonAmnt	**-0.039***	**0.067***	**-0.114***	0.023
ZIP Code Level					
Median Age (in Years)	*Intercept*	0.007	**-0.013***	**-0.044***	0.017
	SolConDur	0.004	-0.001	-0.014	0.013
	Ln(SolConDur)	-0.017	0.001	0.058	-0.023
	ConConDur	**-0.005***	-0.001	-0.006	-0.005
	Ln(ConConDur)	**0.038***	0.012	0.069	0.079
	DiscDonAmnt	0.042	-0.015	-0.128	0.067
Percentage of White	*Intercept*	0.001	-0.001	0.004	0.001
	SolConDur	-0.001	**0.002***	-0.002	-0.007
	Ln(SolConDur)	-0.001	**-0.008***	0.010	0.031
	ConConDur	**-0.001***	0.000	-0.002	-0.002
	Ln(ConConDur)	**0.015***	-0.001	0.039	0.032
	DiscDonAmnt	0.015	-0.002	0.016	0.016
Median Years of Schooling	*Intercept*	-0.053	0.061	**-0.510***	**-0.554***
	SolConDur	0.015	0.060	**0.323***	-0.136
	Ln(SolConDur)	0.010	-0.123	**-1.086***	0.478
	ConConDur	-0.018	-0.011	0.043	-0.058
	Ln(ConConDur)	0.208	0.164	-1.045	0.914
	DiscDonAmnt	-0.134	0.192	**-3.242***	-1.239
Persons per Household	*Intercept*	-0.113	0.192	**-2.278***	-0.275
	SolConDur	-0.128	-0.033	-0.181	-0.604
	Ln(SolConDur)	0.240	0.052	0.175	2.655
	ConConDur	**-0.151***	0.003	**-0.418***	-0.059
	Ln(ConConDur)	**1.305***	0.115	**7.370***	1.420
	DiscDonAmnt	1.113	0.325	-0.358	-1.486
Percentage of Households which are Families	*Intercept*	0.005	-0.001	**0.074***	0.002
	SolConDur	0.000	0.003	0.016	0.013
	Ln(SolConDur)	0.005	-0.008	-0.050	-0.057
	ConConDur	**0.006***	0.000	**0.016***	0.000
	Ln(ConConDur)	**-0.053***	-0.004	**-0.303***	-0.025
	DiscDonAmnt	-0.036	0.003	-0.073	0.006

Estimate Median Income (in 1000 $)	*Intercept*	**-0.013***	-0.003	-0.003	0.015
	SolConDur	-0.004	-0.001	-0.012	0.007
	Ln(SolConDur)	0.013	0.001	0.036	-0.022
	ConConDur	-0.001	-0.001	**-0.007***	0.002
	Ln(ConConDur)	0.005	0.005	**0.128***	-0.034
	DiscDonAmnt	**-0.039***	-0.010	**0.124***	0.024
Median Home Value (in 1000 $)	*Intercept*	**0.003***	**0.002***	0.005	0.004
	SolConDur	0.001	0.000	0.000	-0.002
	Ln(SolConDur)	-0.005	0.001	-0.001	0.002
	ConConDur	**0.001***	0.000	0.001	0.001
	Ln(ConConDur)	**-0.010***	-0.004	-0.021	-0.012
	DiscDonAmnt	0.011	0.001	0.009	0.003
Total No. of Tradelines	*Intercept*	**0.088***	-0.018	-0.054	0.115
	SolConDur	0.017	**-0.029***	**-0.113***	0.045
	Ln(SolConDur)	-0.049	0.072	**0.329***	-0.011
	ConConDur	-0.005	0.002	0.025	0.003
	Ln(ConConDur)	0.070	-0.027	-0.263	0.022
	DiscDonAmnt	**0.383***	-0.018	-0.025	0.213
No. of Protestant Churches per Capita	*Intercept*	0.604	-0.808	-0.841	0.022
	SolConDur	-0.074	-0.093	-0.101	0.030
	Ln(SolConDur)	-0.337	0.035	-0.279	0.041
	ConConDur	-0.001	0.054	0.188	-0.012
	Ln(ConConDur)	0.021	-0.338	0.767	0.215
	DiscDonAmnt	0.075	0.237	-1.311	0.046
No. of Non-Protestant Churches per Capita	*Intercept*	0.088	-0.272	0.107	0.327
	SolConDur	-0.126	-0.046	-0.001	0.343
	Ln(SolConDur)	0.087	0.211	0.003	0.016
	ConConDur	0.116	0.075	0.052	-0.040
	Ln(ConConDur)	0.105	0.138	-0.161	-0.217
	DiscDonAmnt	0.365	0.208	-0.021	-0.114
Average Rating of Schools	*Intercept*	-0.070	-0.007	0.026	0.104
	SolConDur	0.045	0.008	0.107	0.043
	Ln(SolConDur)	-0.265	0.057	-0.349	-0.306
	ConConDur	0.018	-0.005	-0.051	0.007
	Ln(ConConDur)	**-0.314***	0.042	0.499	0.060
	DiscDonAmnt	**-0.444***	-0.145	0.757	0.901
County Level					
Percentage of Republican Votes	*Intercept*	0.001	0.001	-0.014	-0.007
	SolConDur	-0.004	**-0.004***	0.009	-0.003
	Ln(SolConDur)	0.011	**0.013***	**-0.049***	0.004
	ConConDur	0.000	0.000	0.004	0.001
	Ln(ConConDur)	0.003	0.007	**-0.060***	-0.006
	DiscDonAmnt	0.006	0.017	**-0.074***	-0.017
Total Crime per Capita	*Intercept*	**1.185***	0.557	-2.534	-2.270
	SolConDur	-0.222	0.200	1.479	-1.308
	Ln(SolConDur)	0.726	-0.202	-5.916	-0.814
	ConConDur	**0.369***	-0.006	0.302	0.077
	Ln(ConConDur)	**-2.981***	0.065	3.100	0.492
	DiscDonAmnt	**3.393***	1.564	-6.678	-0.647

Table 8a. Significant Effects for Observed Heterogeneity in Donation Behavior with respect to Community Characteristics Groups

Donation Program 1		*Donation Program 2*	
Incidence	*Amount*	*Incidence*	*Amount*
Demographic Composition			
Fraction of Male No. of Lifetime Solicitations Median Age (in Years) Percentage of White Persons per Household Percentage of Households with Families	Fraction of Male No. of Lifetime Solicitations Median Age (in Years) Percentage of White	Fraction of Male No. of Lifetime Solicitations Median Age (in Years) Median Years of Schooling Persons per Household Percentage of Households with Families	No. of Lifetime Solicitations Median Years of Schooling
Financial Capability			
Estimated Median Income (in 1000$) Median Home Value (in 1000$) Total No. of tradelines	Median Home Value (in 1000$) Total No. of Tradelines	Estimated Median Income (in 1000$) Total No. of Tradelines	
Educational Environment			
Average Rating of Schools			
Religious Inclinations			
Political Ideology			
	Percentage of Republican Votes	Percentage of Republican Votes	
Public Safety			
Total Crime per Capita			

Table 9. Predictive Performance using Mean-Absolute Deviation for Donation Program 1

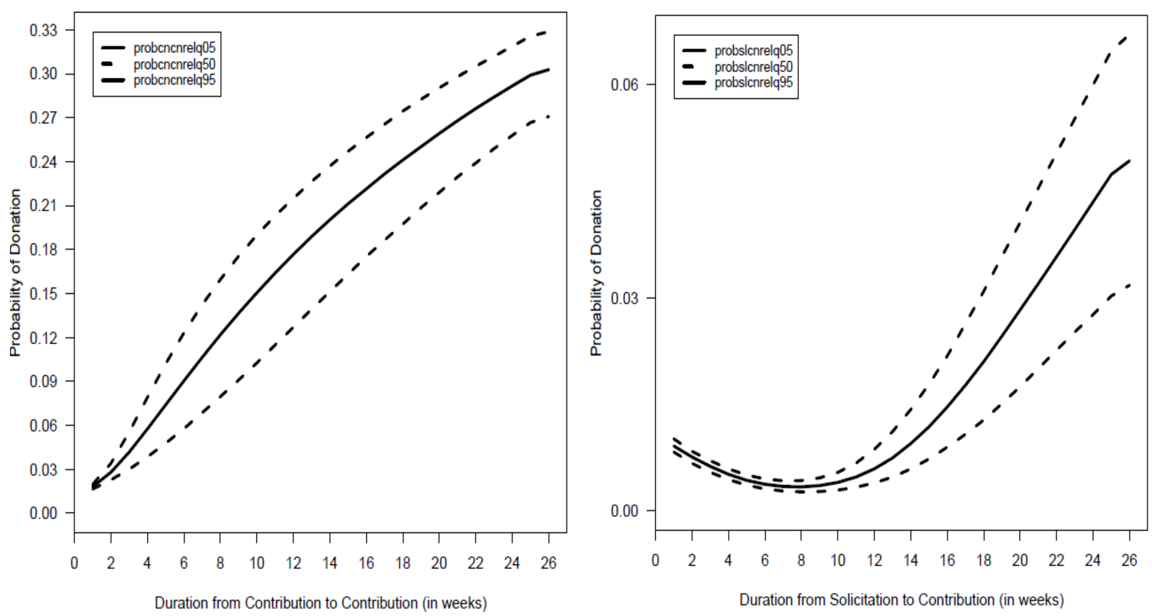
<table>
<tr><td>Donation Program 1</td><td colspan="5">Amount of Contributions</td></tr>
<tr><td rowspan="7">Frequency of Contributions</td><td></td><td colspan="2">Low</td><td colspan="2">High</td></tr>
<tr><td rowspan="3">Low</td><td colspan="2">Donor Segment A</td><td colspan="2">Donor Segment B</td></tr>
<tr><td rowspan="2">Observed
335</td><td>Predicted
461</td><td rowspan="2">Observed
256</td><td>Predicted
96</td></tr>
<tr><td>Correct Predictions
172
(51%)</td><td>Correct Predictions
32
(13%)</td></tr>
<tr><td rowspan="3">High</td><td colspan="2">Donor Segment C</td><td colspan="2">Donor Segment D</td></tr>
<tr><td rowspan="2">Observed
374</td><td>Predicted
207</td><td rowspan="2">Observed
165</td><td>Predicted
366</td></tr>
<tr><td>Correct Predictions
66
(18%)</td><td>Correct Predictions
71
(43%)</td></tr>
</table>

Table 10. How do Coupon Targeting and Face-values affect Brand Purchase Behavior?

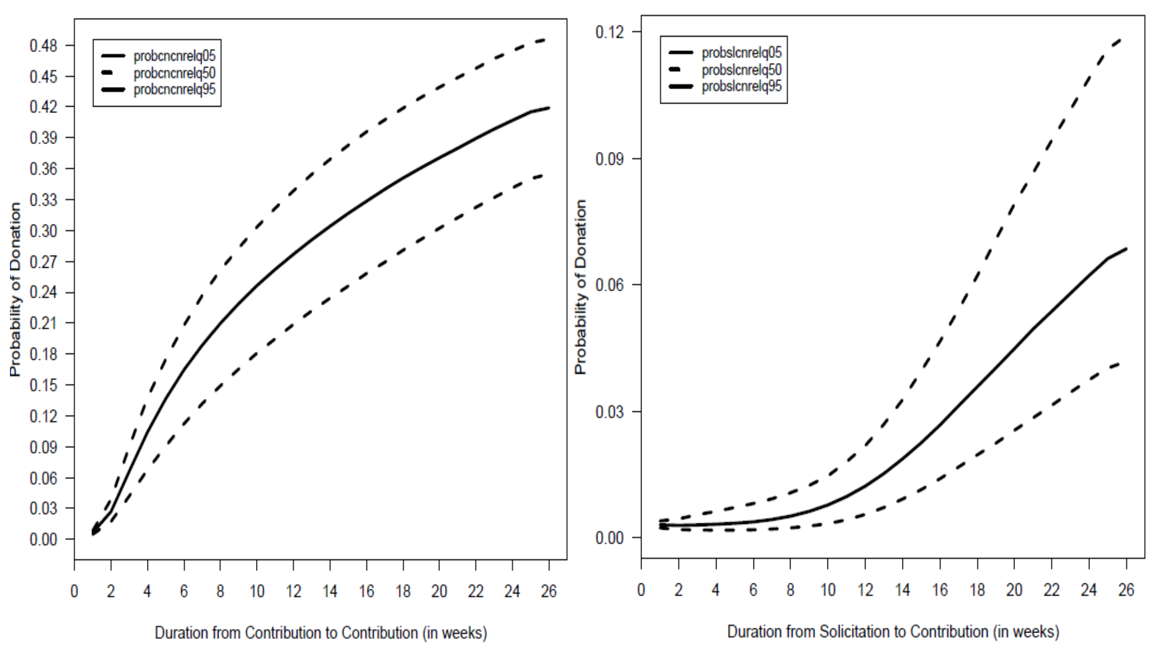
Redemption of Coupon → Offering of Coupon ↓	**No** ($f_{ijt} = 0$)	**Yes** ($f_{ijt} = 1$)
No ($q_{ijt} = 0$)	$\phi_{ij} p_{ijt}$	
Yes ($q_{ijt} = 1$)	$(\phi_{ij} p_{ijt} + \mu_{ij})$	$(\phi_{ij} p_{ijt} + \mu_{ij} - \phi_{ij} f_{ijt})$

FIGURES

Figure 1. Marginal Effects of Duration Variables (from Solicitation and Previous Donation) on Contribution Incidence and Corresponding Reliability Estimates (5th and 95th percentile)



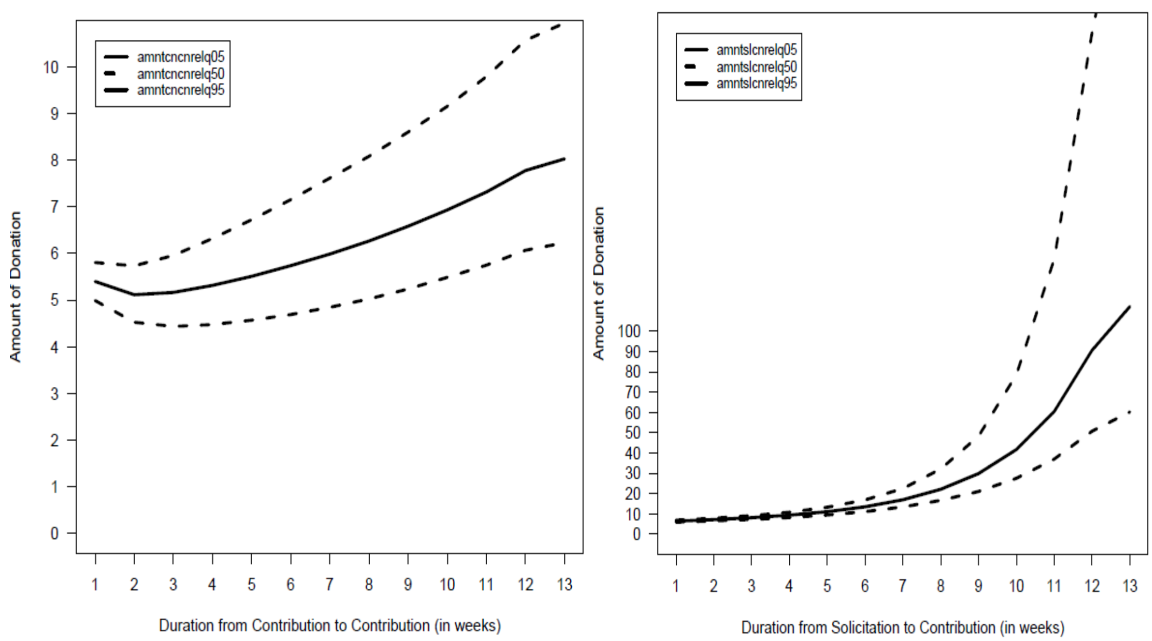
Program 1



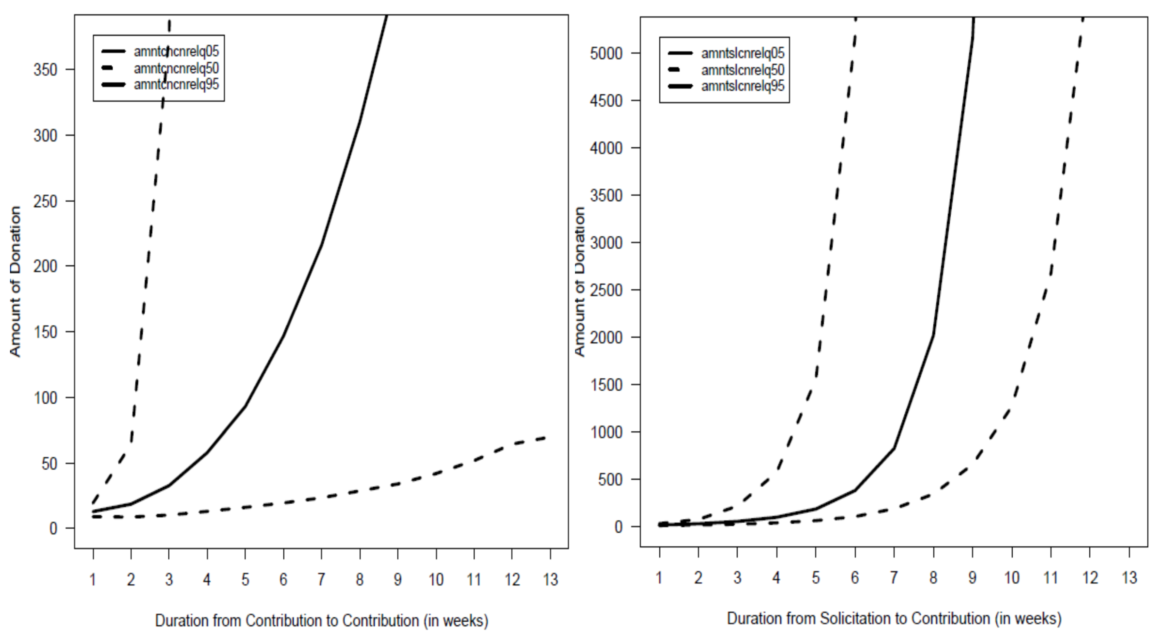
Program 2

Note: In the figures above, 'q5' denotes 5th percentile and 'q95' denotes 95th percentile

Figure 2. Effect Marginal Effects of Duration Variables (from Solicitation and Previous Donation) on Contribution Amount and Corresponding Reliability Estimates (5th and 95th percentile)



Program 1



Program 2

Note: In the figures above, 'q5' denotes 5th percentile and 'q95' denotes 95th percentile

Figure 3. How Learning from Brand Purchase Behavior can be Valuable for Retailer in Targeting?

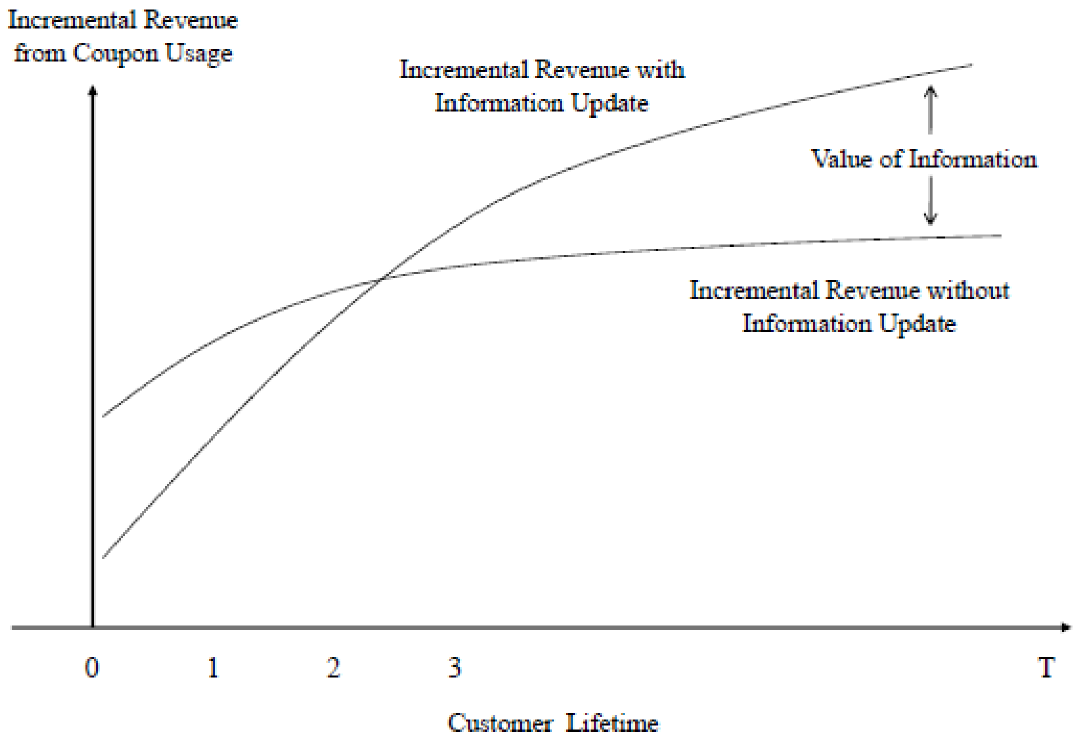
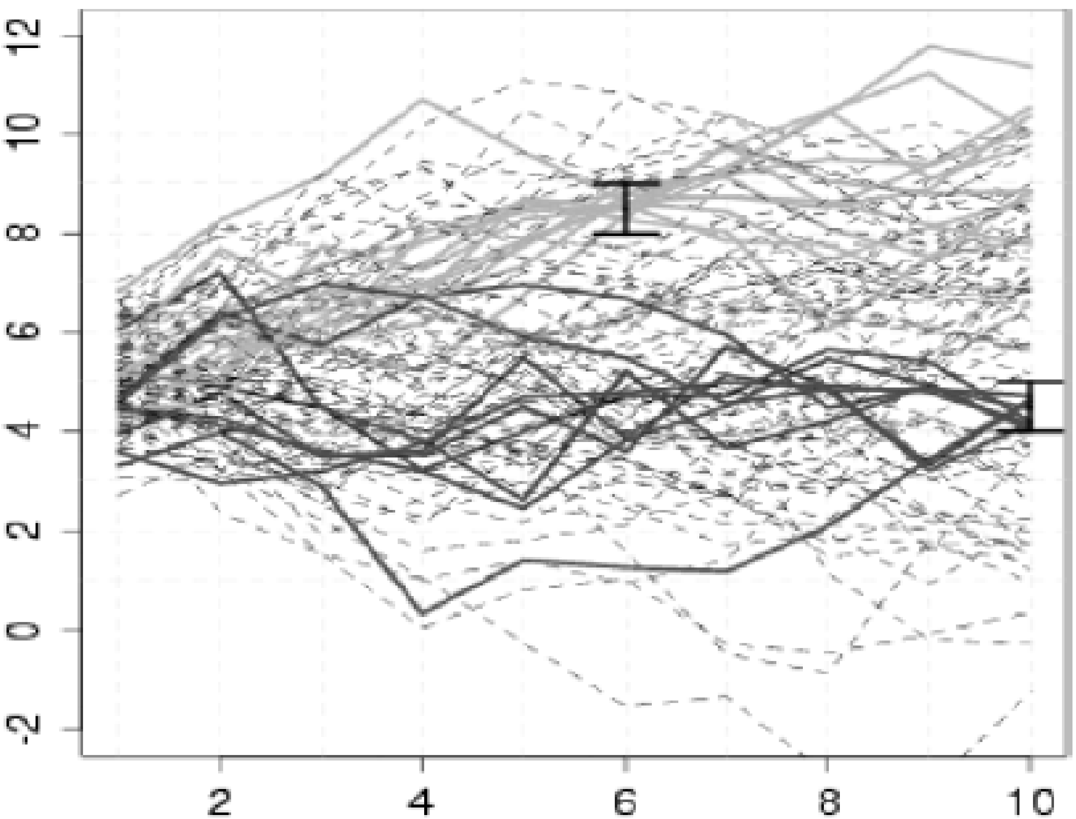


Figure 4. Sequential Decision Making using Backward Induction: Grid-Based Optimization over Constrained Action Space



REFERENCES

Allenby, Greg M., Robert P. Leone and Lichung Jen (1999), “A Dynamic Model of Purchase Timing with Application to Direct Marketing,” Journal of the American Statistical Association, 94(446), 365-374.

Allenby, Greg M. and Peter E. Rossi, “Marketing Models of Consumer Heterogeneity,” Journal of Econometrics, 89, 57-78.

Armstrong. J. Scott and Roderick Brodie (2004), “Effects of Portfolio Planning Methods on Decision Making: Experimental Results,” International Journal of Research in Marketing, 11(1). 73-84.

Baesens, Bart, Stijn Viaene, Dirk Van den Poel, Jan Vanthienen and Guido Dedene (2002), “Bayesian Neural Network Learning for Repeat Purchase Modeling of Direct Marketing,” European Journal of Operational Research, 138(1), 191-211.

Basu, Amiya K., Atasi Basu and Rajeev Batra (1995), “Modeling the Response Pattern to Direct Marketing Campaigns,” Journal of Marketing Research, 32(2), 204-212.

Bellair, Paul E. (1997), “Social Interaction and Community Crime: Examining the Importance of Neighbor Networks,” Criminology, 35(4), 677-704.

Bennett, Roger (2003), “Factors Underlying the Inclination to Donate to Particular Types of Charity,” International Journal of Nonprofit and Voluntary Sector Marketing, 8(1), 12-29.

Bitran, Gabriel R. and Susana V. Mondschein (1996), “Mailing Decisions in the Catalog Sales Industry,” Management Science, 42(9), 1364-1381.

Bucklin, Randolph E. and Sunil Gupta (1992), “Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach,” Journal of Marketing Research, XXIX, 201-215.

Bult, Jan R. and Tom Wansbeek (1995), “Optimal Selection for Direct Mail,” Marketing Science, 14(4), 378–394.

Bussell, Helen and Deborah Forbes (2002), “Understanding the Volunteer Market: The What, Where, Who and Why of Volunteering,” International Journal of Nonprofit and Voluntary Sector Marketing, 7(3), 244-257.

DeMarzo, Peter M., Ron Kaniel and Ilan Kremer (2004), “Diversification as a Public Good: Community Effects in Portfolio Choice,” The Journal of Finance, 59(4), 1677-1716.

Diepen, Merel Van, Bas Donkers and Philip Hans Franses (2009), “Dynamic and Competitive Effects of Direct Mailings: A Charitable Giving Application,” Journal of Marketing Research, XLVI, 120-133.

Elsner, R., M. Krafft and A. Huchzermeier (2004), “Optimizing Rhenania’s Direct Marketing Business Through Dynamic Multilevel Modeling (DMLM) in a Multicatalog-Brand Environment,” Marketing Science, 23(2) 192–206.

Fader, Peter S. and Bruce G.S. Hardie (1996), "Modeling Consumer Choice Among SKUs," Journal of Marketing Research, 33 (November), 442-452.

Fader, Peter S., Bruce G. S. Hardie and Ka Lok Lee (2005), “”Counting Your Customer” the Easy Way: An Alternative to the Pareto/ NBD Model,” Marketing Science, 24(2), 275-284.

Fader, Peter S., Bruce G. S. Hardie and Jen Shang (2010), “Customer-Base Analysis in a Discrete-Time Noncontractual Setting,” Marketing Science, forthcoming.

Frank, R. E., W. F. Massy and Y Wind (1972), “Market Segmentation,” Prentice-Hall, Engelwood Cliffs.

Gill, Jeff (2008), “Bayesian Methods: A Social and Behavioral Sciences Approach,” Chapman and Hall, Boca Raton, Florida.

Gonul, F. F. and Mengze Shi (1998), “Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models,” Management Science, 44(9), 1249–1262.

Gonul, Fusun F. and Frenkel Ter Hofstede (2006), “How to Compute Optimal Catalog Mailing Decisions,” Marketing Science, 25(1), 65-74.

Grewal, Dhruv, Gopalkrishnan R. Iyer, Wagner A. Kamakura, Anuj Mehrotra and Arun Sharma (2008), “Evaluation of Subsidiary Marketing Performance: Combining Process and Outcome Performance Metrics,” Journal of the Academy of Marketing Science, 37(2), 117-129.

Harvey, James W. (1990), “Benefit Segmentation for Fund Raisers,” Journal of the Academy of Marketing Science, 18(1), 77-86.

Hogan, Dennis P. and Evelyn M. Kitagawa (1985), “The Impact of Social Status, Family Structure, and Neighborhood on the Fertility of Black Adolescents,” The American Journal of Sociology, 90(4), 825-855.

Ibrahim, Joseph G., Min-Hui Chen and Debajyoti Sinha (2010), “Bayesian Survival Analysis,” Springer Series in Statistics.

Jain, Dipak C., Naufel Vilcassim and Pradeep Chintagunta (1994), “A Random Coefficients Logit Brand Choice Model Applied to Panel Data,” Journal of Business & Economic Statistics. 12(3): 317-328.

Jones, Andrew and John Posnett (1991), “Charitable Donations by UK Households: Evidence from the Family Expenditure Survey,” Applied Economics, 23, 343 351.

Kamakura, Wagner A. and Gary J. Russell (1989), “A Probabilistic Choice Model for Market Segmentation and Elasticity Structure,” Journal of Marketing Research, XXIV, 379-390.

Kitchen, Harry and Richard Dalton (1990), “Determinants of Charitable Donations by Families in Canada: A Regional Analysis,” Applied Economics, 22, 285-299.

Leclere, Felicia B., Richard G. Rogers and Kimberley Peters (1998), “Neighborhood Social Context and Racial Differences in Women’s Heart Disease Mortality,” Journal of Health and Social Behavior, 39(2), 91-107.

Manchanda, Puneet, Asim Ansari and Sunil Gupta, “The “Shopping Basket”: A Model for Multicategory Purchase Incidence Decisions,” Marketing Science, 18(2), 95-114.

Morrison, Donald G. (1969), “On the Interpretation of Discriminant Analysis,” Journal of Marketing Research, VI, 156-163.

Muller, Peter, Don A. Berry, Andy P. Grieve, Michael Smith and Michael Krams (2007), “Simulation-based Sequential Bayesian Design,” Journal of Statistical Planning and Inference, 137(10), 3140-3150.

Narendranathan, W. and M. B. Stewart (1993), “How Does the Benefit Effect Vary as Unemployment Spells Lengthen?” Journal of Applied Econometrics, 8(4), 361-381.

Peeples, Faith and Rolf Loeber (1994), “Do Individual Factors and Neighborhood Context Explain Ethnic Differences in Juvenile Delinquency?,” Journal of Quantitative Criminology, 10(2), 141-157.

Piersma, Nanda and Jedid-Jah Jonker (2004), “Determining the Optimal Direct Mailing Frequency,” European Journal of Operational Research, 158(1), 173-182.

Reinartz, Werner J. and V. Kumar (2003), “The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration,” Journal of Marketing, 67(1), 77 99.

Reinartz, Werner, Jacquelyn S. Thomas and V. Kumar (2005), "Balancing Acquisition and Retention Resources to Maximize Customer Profitability," Journal of Marketing, 69, 63-79.

Ridgeway, Greg and David Madigan (2003), "A Sequential Monte Carlo Method for Bayesian Analysis of Massive Datasets," Data Mining and Knowledge Discovery, 7(3), 301-319.

Rossi, Peter E., Robert E. McCulloch and Greg M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," Marketing Science, 15(4), 321-340.

Rossi, Peter E., Robert E. McCulloch and Greg M. Allenby (2005), "Bayesian Statistics and Marketing," Wiley Series in Probability and Statistics.

Rust, Ronald T. and Peter C. Verhoef (2005), "Optimizing the Marketing Interventions Mix in Intermediate-Term CRM," Marketing Science, 24(3), 477-489.

Sampson, Robert, J., Jeffrey D. Morenoff and Thomas Gannon-Rowley (2002), "Assessing 'Neighborhood Effects': Social Processes and New Directions in Research," Annual Review of Sociology, 28, 443-478.

Sargeant, Adrian (1999), "Charitable Giving: Towards a Model of Donor Behavior," Journal of Marketing Management, 15, 215-238.

Schlegelmilch, Bodo B., Adamantios Diamantopoulos and Alex Love (1997), "Characteristics Affecting Charitable Donations: Empirical Evidence from Britain," Journal of Marketing Practice: Applied Marketing Science, 3(1), 14-28.

Schlegelmilch, Bodo B., Alex Love and Adamantios Diamantopoulos (1997), "Responses to Different Charity Appeals: The Impact of Donor Characteristics on the Amount of Donations." European Journal of Marketing, 31(8), 548-560.

Schlegelmilch, B. B. and A. C. Tynan (1989), "The Scope of Market Segmentation Within the Charity Market," an Empirical Analysis," Managerial and Decision Economics, 10, 127-134.

Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," Marketing Science, 13(1), 41-67.

Shannon, C.E. (1948), "A Mathematical Theory of Communication," Bell System Technical Journal, 27, 379–423 & 623–656.

Simester, Duncan I., Peng Sun and John N. Tsitsiklis (2006), "Dynamic Catalog Mailing Policies," Management Science, 52(5), 683-696.

Venkatesan, Rajkumar, V. Kumar and Timothy R. Bohling (2007), "Optimal CRM using Bayesian Decision Theory: An application for Customer Selection," Journal of Marketing Research, 44 (4), 579 – 594.

Verdinelli, Isabella and Joseph B. Kadane (1992), "Bayesian Designs for Maximizing Information and Outcome," Journal of the American Statistical Association, 87(418), 510-515.

Wedel, Michel and Wagner A. Kamakura (2002), "Market Segmentation: Conceptual and Methodological Foundations," Kluwer Academic Publishers, Boston.

Wind, Yoram, Vijay Mahajan and Donald J. Swire (1983), "An Empirical Comparison of Standardized Portfolio Models," Journal of Marketing, 47, 89-99.

VITA

Shameek Sinha was born in Calcutta, India on September 2, 1979, the son of Madhuchhanda and Utpal Sinha. He completed his primary school at Nava Nalanda, his middle school at Ramakrishna Mission Boys' High School and high school at St. Xavier's College, Calcutta (under University of Calcutta). He chose economics, statistics and mathematics as his specializations in high school. Shameek entered Presidency College, Calcutta (under University of Calcutta) for his undergraduate studies and graduated with Bachelor of Science (Honors) in Economics. He then pursued his graduate studies at the Indian Statistical Institute (New Delhi) and obtained a Master of Science in Quantitative Economics. After completing his graduate studies in India, Shameek entered the Doctoral program in Economics at the University of Texas at Austin in the Fall of 2003. He obtained a Master of Science in Economics before moving to the Doctoral program in Marketing at the Red McCombs School of Business at the University of Texas at Austin in the Fall of 2006. In the Marketing doctoral program , he received a Master of Science in Marketing as well.

Permanent Address: 1003 Justin Lane, Argosy at Crestview Apartment, Apt. 1033, Austin, TX – 78757-2644, USA

This dissertation was typed by Shameek Sinha