Essays on Customization Applications in Marketing

Adiguzel, Feray

Chapter 5
Conclusion and Discussion

5.1 Introduction

The primary objective of this thesis is to develop and validate new methodologies to improve the collection of data and the effectiveness of promotion customization. For this purpose, we use the Bayesian approach in every stage of problem solving, inference, estimation and decision making. This thesis contains two essays. The first essay deals with how to improve collection of data. We recommend using split questionnaires for long questionnaires, common in marketing, and develop a methodology to design optimal split questionnaires. The second essay is a cross-category promotion customization problem. We fit and estimate a model across multiple categories, which allows cross-category promotion strategies. We customize optimal promotion design with a combinatorial optimization approach. In this chapter, the main conclusions are summarized and further research is suggested.

5.2 Summary and Conclusions

In the first chapter, we explained the term “customerization”. Marketing managers need new online marketing strategies such as new methods of interacting with customers because of the migration of marketing to the online environment. Customization and customerization are two different concepts that should not be confused (Wind and Rangaswamy, 2001). Customerization indicates the customer’s individual likes and dislikes which are placed at the center of every stage of the marketing process, rather than only tailoring the offering. Customerization can be summarized as merging strategies of one-to-one marketing, personalization, targeting and mass-customization. Successful customerization strategies should combine supply and demand sides. From this perspective, the possibilities of customizing marketing mix instruments from the seller and buyer side are explained in Chapter 1. To develop new customerization strategies, we need customer information (likes, dislikes, lifestyles, purchase habits, some background variables etc.), which is critical for identifying, differentiating, and interacting with customers. This information can be collected with questionnaires. Better tools for data collection and better models for customization will help managers or market researchers make better decisions. The Bayesian framework will prove to be useful. Two applications are presented on collecting consumer data on soft variables such as lifestyles or consumer satisfaction in Chapter 3 and cross-category promotions in Chapter 4.

In the second chapter, we present why the Bayesian approach is particularly appropriate to the decision orientations of marketing problems. In the Bayesian approach, all available information is used to reduce the amount of uncertainty which is present in an inferential or decision-making problem. The main reason for the increased usage of Bayesian methods in marketing in the last decade is not only the increasing capacity of computers and the success of MCMC algorithms to solve complex marketing problems, but also the reliance on the characteristics of marketing data, the necessity to approach marketing problems as a decision problem and the flexibility and robustness of Bayesian methods. Marketing models with latent variables, missing data, mixed outcome data, heterogeneity of coefficients, nonlinearity, discrete data and more, are easy to estimate with in the Bayesian framework. Since the Bayesian paradigm uses all information and merges prior information with observed data to estimate models (updating information), the Bayesian paradigm is optimal for decision problems in marketing. The Bayesian decision process considers the estimation or model uncertainty in the complete process of problem solving.

In the third chapter, we focus on split questionnaires to collect data instead of using the more typical long questionnaires (i.e. more than 20 minutes) in marketing, since they offer the potential to obtain higher quality information from respondents faster and at a substantially lower cost. In split questionnaires, different respondents respond different parts of the questionnaire. That is, we have different versions of the questionnaire that are shorter than the whole questionnaire. After generating different split questionnaires and administrating them to respondents, we impute data for the missing parts using the other people's responses to those missing parts. In the end, we obtain almost the same information with split questionnaires as with complete lengthy questionnaires, but in a shorter time with less cost and obtaining better quality responses (less item nonresponse, higher response and more accurate responses). We propose a methodology to generate split questionnaire versions based on some prior information and for this we use optimal experimental design methods. We generate many designs with the modified Federov algorithm to search over the design space using the Kullback-Leibler distance as a design criterion, and illustrate that good designs are feasible. We present synthetic data results for algorithm performance, real data results for statistical efficiency (i.e. we show that we obtain almost the same information with split questionnaires compared to full questionnaires) and field study results that reveal behavioral efficiency. The statistical and behavioral efficiency of split questionnaire designs are shown by comparing them to full questionnaires or questionnaires constructed with ad-hoc methods.

In the fourth chapter, the cross-category promotion design problem is presented. Currently, many multicategory models are restricted to a smaller number of categories, and the main purpose in these studies is to understand any type of demand relationship across product categories (substitution, complementarity or independence). Retailers can use cross-category relatedness for delivery of point-of-purchase materials, cross-category coupons, creative store layout, and online feature ad design. In marketing, we need better models to understand consumers' multicategory preferences. In our application, we consider not only category interdependencies, but also purchase incidence and expenditure interdependencies using a hierarchical Bayes type-2 tobit model. We approach the problem of optimally promoting a limited set of categories as a combinatorial optimization problem and use a design generating algorithm to generate many promotion designs to find the optimal one. The objective function is the maximum profit change if the category is promoted for the selected categories. The approach presented here can be applied to other possible marketing offerings, which can be a product, service, combination of product and service, or a bundle of products and/or services. During the model estimation and optimization, we use a Bayesian approach which allows considering estimation uncertainty and parameter uncertainty. In decision processes in marketing, the degree of uncertainty in making decisions needs to be communicated effectively to managers. As we know, managers are generally risk-averse, and analysis involving exogenous variables can generate predictions that vary greatly in their precision. From this perspective, predictive uncertainty should be communicated to managers, and thus they can favor decisions that correspond to more certain predictions, or they can collect additional information to further reduce uncertainty.

5.3 Limitations and Future Research

Despite the many developments that have already taken place in conjoint questionnaire design literature, there is a need to collect better quality data with long questionnaires in surveys in marketing. Successful customization strategies depend on better models and better optimization algorithms. We propose a combinatorial optimization approach, which allows optimal assignment of promotions across multiple categories. In this section, we consider those avenues, as well as limitations and possible (or planned) extensions that are related to the research presented in Chapter 3 and 4.

5.3.1 Split Questionnaires

First, we present some methodological limitations. We assume a multivariate normal distribution for variables in questionnaire in Chapter 3. The SQD method could be extended to accommodate binomial data or mixtures of categorical and continuous data, based on the general location-scale model (Olkin, 1961) to enable one to optimally split and impute questionnaires in a wider variety of questionnaire design problems. Although we mention how we can extend the proposed questionnaire design problem to the mixed data case with the general location model, we have not included results into this thesis. The extension of the proposed method in design and imputation stage for the mixed data case is in progress.

One limitation of this research could be our multiple imputations based on the multivariate normal distribution. We intend to extend the imputation procedure to the mixed data case, on which there is already existing literature. Although we imputed data without considering individual heterogeneity, there are some studies that consider multiple imputation at the individual level. Gelman et al. (1998) propose a multiple imputation procedure, which assumes an imputation model that allows one to include covariates on the individual and the survey. In this procedure, the individual heterogeneity enters into the imputation procedure through the multivariate normal data model, with a common covariance matrix and differing mean vectors concerning the survey levels at the individual level.

We used a greedy algorithm to design within block designs, however the sensitivity of a within block design to prior information is relatively high with this algorithm. One would prefer to generate those designs with different algorithms that reduce the sensitivity of this design to prior information. Therefore, better design algorithms for the within block designs are an avenue for future research.

One of the problems of using prior information to construct split designs is that we are still using a full questionnaire from a pilot study or sub sample. Therefore, any undesirable response styles from a full questionnaire may still affect our generation of split questionnaire designs. However, we can prevent this problem by checking the response pattern of subjects in a full questionnaire. We can eliminate or correct the questions with bad response styles from a full questionnaire, or we can develop more extensive models that take response styles into account.

It would also be of future interest to develop dynamic algorithms to optimize questionnaire designs. SQD methods are arguably important tools in web-based surveys, online panel surveys and pop-up questionnaires (Comley, 2000). Here, respondent burden is an even more important issue, and our method could be extended to allow dynamic updating of the split questionnaire design for each respondent as more data comes in. Using some past information from subjects, one can then very quickly and efficiently customize the split questionnaire to individual customers. We believe for “real time” marketing decisions, online questionnaires can be an important tool in the future. The main limitation of an online survey is its length. Online surveys should be short. Fram and Grady (1995) found consumers unwilling to respond to lengthy surveys administered online. Principals at NFO Research, Inc. also report that participation rates drop dramatically when online surveys become long (e.g., more than 40 items). As a result, questionnaire constructs and concepts must be captured parsimoniously in interactive surveys, and split questionnaires methodology can be applied to design online surveys. In fact, online questionnaires already have become a main interest in conjoint analysis. Since these questionnaires can be used for pricing and new product development analysis. Sometimes one may need to merge survey data and behavior data. For example, De Bruyn et al. (2005) did a study on online conjoint questionnaires, in which they used survey data questions and behavioral data to collect consumer preferences using shorter conjoint questionnaires. Modifications of our method to “real-time” decisions are also possible. For instance, we can design split questionnaires in two stages. First, we may need to segment customers with certain questions (or common questions) at an initial stage, and then customize questionnaires based on these segments to collect data more efficiently in the second stage. We may have some prior information to understand which questions can be used to classify respondents to segments. When respondents start to respond the questions, this prior information can be updated at the second stage using the Kullback-Leibler distance as an information statistic at each stage. The order of the question can then be decided for each individual based on this information statistic. At any moment during the questionnaire, we can know the probability of the respondent belonging to each of the segments using the finite mixture model (Kamakura and Wedel, 1998).

Another possibility for future research is to study the optimal sample size for each version of the split questionnaire. Although we distributed each distinct split (versions of the questionnaire) to respondents evenly in our application, “sample size” for each questionnaire is an important issue one should consider. Each split should be distributed to a sufficient number of respondents for validity. Then the question becomes how can we estimate efficient sample sizes for each split? Are there some versions of the questionnaire that need more subjects, while others do not? Should sample size depend on the number of questions in the questionnaire, or on the information content of each split questionnaire? We think these questions are of interest for future research.

Although we do not need common questions to design split questionnaires, using them may increase the efficiency of imputations. In split questionnaire design where certain common questions are contained in both versions of the questionnaire, attention should be given to the ordering and positioning of these common questions to reduce potential response bias and/or carryover effect. For population surveys conducted regularly over time, variables such as gender, age, background, etc. can easily be conceived as common variables when the change of population dynamics over a certain time period can be ignored. If it is necessary to ask certain questions to every respondent, we may use them as common variables.

In the future, we probably will see more applications of split questionnaires in media and purchasing behavior panel surveys. Rassler (2002) claims that questionnaires used in the television measurement panel and the purchasing behavior panel can be reorganized to create suitable blocks of variables. Currently used methods for this purpose, such as data fusion, depend on the conditional independence assumption and therefore suffer from identification problems. On the other hand, by overlapping blocks of questions, split questionnaires can provide a solution to the identification problem by overcoming the conditional independence assumption.

There are several issues to be resolved in future research. We generated SQDs using the number of splits as an external constraint. SQDs can be easily generated under different constraints that arise in practical applications, but the performance of the SQD under such constraints remains to be investigated. As an illustration, we investigated the effect of the constraint of five blocks for each split in our empirical application. We obtain a much larger reduction of the number of questions with the five-block constraint, however, at the same time, the percentage of missing information increases and the performance of the SQD decreases in an absolute sense (although its performance relative to the RQD seems to improve). Our purpose is to eliminate questions at a minimum cost of information loss. Therefore, considering all possible splits without any constraint may be more desirable, since there is more opportunity to borrow information between blocks, which increases the efficiency of the imputations. Nevertheless, in applications, additional constraints, such as the number of blocks of questions (and indirectly the total number of questions to ask in a survey) in each split of the questionnaire may be important. These kinds of constraints can ensure less respondent fatigue, which is very useful for practitioners. In short, the pay-off of reducing respondent burden versus information loss is another important topic for future research.

5.3.2 Customization

First, we discuss a few methodological issues. The main limitation in the current approach is the design generating algorithm used, since it is limited to twenty categories. As the number of categories becomes large, this approach will become infeasible. We can possibly increase this number using different design generating algorithms to find optimal product promotions. Additionally, in case we have many explanatory variables, we may consider estimating the error covariance of individual level marketing mix parameter estimates with a factor approach. We can model the large covariance matrix using a parsimonious factor analytic model.

There are a number of limitations of the model used in Chapter 4, all of which provide new directions for future research. The model developed ignores brand choice. We may need a better model to accommodate households' brand choice decisions within each product category. For that, we can use the multinomial logit (or probit) model for households' conditional brand choice within each product category, in combination with the multivariate probit model for category purchase incidence. After including brand choice in the category expenditure and incidence model, we should also consider it in the optimization problem. Then our design problem can be defined as choosing a certain number of categories to promote from among many, and from these categories, which brand should be selected for a sales promotion. There are two possible approaches for the brand promotion allocation problem. The first one is to use a two-stage design generating algorithm. At the first stage, we can decide which categories to promote, and at the second stage, we can choose a brand to promote from the selected categories. The second approach is that we can select a brand from each category using the greedy algorithm that we also used to design within block designs.

Our model can be improved by including the consumer's budget. We expect a high degree of correlation between purchase incidences and the consumer's shopping budget. In fact, budget constraints may cause cross-category dependencies as well. However, the budget constraint is unobserved. Assuming a utility function for each category, we may assume that a customer chooses to allocate his/her dollars across categories by selecting that allocation that maximizes this utility function. Budget constraints in the context of searching for cross-category effects were previously applied by Song and Chintagunta (2003).

Another direction would be to include dynamics (i.e. state dependence) into price and promotion effects across categories. Then, the allocation of promotion varies over time and depends on the reactions to the categories promoted last. State dependence can enter the model with lagged purchase incidence variables, or with time varying marketing mix variables.

There are some studies within the context of random utility models that investigate the effects of a household's current choice on its future choices. Seetharaman, Ainslie and Chintagunta (1999) investigate similarities and differences in household state dependence behavior across multiple categories, relying on the effects of household and category variables. They study cross-category state dependence effects across five categories (ketchup, peanut butter, margarine, toilet tissue, and tuna). They used a multinomial probit model for choice within a category and a Bayesian variance components model (Ainslie and Rossi, 1998) for the covariation of household response parameters across categories. Optimal promotional allocation based on that approach would be dynamic and change over with time.

The composition of the shopping basket is one of the many multicategory choice phenomena that are encountered by consumers. We can use our approach to model consideration formation sets of consumers. Sometimes, due to limited information processing capacity, consumers narrow their options to simplify their decision task. We use the term "consideration set" in marketing, which covers the brands consumers consider acceptable for the next purchase. Explicitly, consideration set refers to the set of brands (a subset of all the brands in the product category) between which a consumer makes an explicit utility comparison or cost benefit trade-off before she makes her brand choice decision. We can extend our approach to include consideration sets. From online transaction data, we observe that in general consumers purchase only a few brands in the category, and sometimes never purchase some of the categories. Each household choice set can be modeled by taking the set of all possible subsets of the available brands which are purchased by the consumer, and assigning a household specific probability distribution on each set (DeSarbo et al., 1995, Chiang et al., 1999). Using this approach, we can generate the set of all possible subsets of brands, using the design generating algorithm for each consumer, based on observed a priori category and brand choices. Such a model would considerably simplify the optimal promotional design problem. Since for each customer, we don't need to use all categories and brands in the category, but only the ones actually considered, our optimization problem can be simplified, and promotion designs can be possible across a large number of categories.

In this thesis, we studied two different topics which can be useful for solving internet marketing problems. Nowadays, the new way of communication to customers is the Internet, and for this reason the importance of developing methodologies in an online environment is increasing. We believe that questionnaire design methods, especially split questionnaire design, will receive more interest in marketing research in the future in order to collect data more efficiently (i.e. cheaper and faster) and with better quality. Moreover, this topic presents many avenues for future research. Customization is an important topic for marketing academics, market researchers and especially marketing managers. In addition to its substantive applications in marketing, there will be increasing opportunities for research on the topic.