MULTICRITERION MARKET SEGMENTATION: A UNIFIED MODEL, IMPLEMENTATION AND EVALUATION

by

Ying Liu

A Dissertation Submitted to the Faculty of the

COMMITTEE OF BUSINESS ADMINISTRATION

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN MANAGEMENT

In the Graduate College

THE UNIVERSITY OF ARIZONA

2007

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation

prepared by Ying Liu

entitled Multicriterion Market Segmentation: a Unified Model, Implementation and Evaluation

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

____ Date: 06/22/2007
Sudha Ram

____ Date: 06/22/2007
Robert Lusch

____ Date: 06/22/2007
Mohan Tanniru

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 06/22/2007
Dissertation Director: Sudha Ram

STATEMENT BY AUTHOR

SIGNED: Ying Liu

ACKNOWLEDGMENTS

I am grateful to my advisor, Dr. Sudha Ram, for her support, encouragements, and expert guidance. She guided and inspired me throughout every state of my research. Dr. Ram spent countless hours reviewing and discussing research related to this dissertation.

My heartfelt thanks and appreciation go out to Dr. Robert Lusch, who was always encouraging with his words of advice. I could not finish this dissertation without his effort to data collection and valuable suggestions.

I am thankful to Dr. Mohan Tanniru for his encouragement and support from the first day of my study in the MIS department.

I am grateful to many friends for their helps and encouragements: Ling Zhu, Dr. Jiannan Wang, Dr. Leon Zhao, Dr. Gang Wang, Dr. Jiexun Li, Dr. Yilu Zhou, Dr. JiaLun Qin, Dr. Ming Lin, Ming Yuan, Zan Huang, Yiwen Zhang and Xin Li.

DEDICATION

To my parents for their love and support

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

LIST OF TABLES

LIST OF TABLES - Continued

LIST OF FIGURES

LIST OF FIGURES - Continued

ABSTRACT

Market segmentation is a multicriterion problem. This dissertation addresses the multicriterion nature of market segmentation with a new unified segmentation model that is derived from a multiobjective conceptual framework. The unified model elegantly solves the intrinsic antagonistic problem of market segmentation by generating a set of Pareto optimal solutions that represent different tradeoffs among multiple conflicting objectives. This dissertation develops an innovative implementation named Multicriterion Market Segmentation using Evolutionary Algorithm (MMSEA). Based on multiobjective evolutionary algorithms, MMSEA overcomes many limitations and disadvantages of existing methods by optimizing multiple objectives simultaneously, searching for globally optimal solutions and generating a set of Pareto optimal solutions. It also suggests the interesting solutions based on the geometric characteristics of Pareto front. The method was applied to customer value and benefit segmentation for the cell phone service market (a descriptive segmentation model) and customer response segmentation for a national retailer (two predictive segmentation models). The empirical evaluation shows that the proposed unified market segmentation model and solution techniques provide the decision makers with many insights and enhanced flexibility that are missing in existing market segmentation methods.

1 INTRODUCTION

Understanding and differentiating customers by their needs and responses to marketing mixes plays a vital role in managing customer relationships. This can be achieved by market segmentation that has been applied in almost every marketing research area including both the consumer and the firm behaviors. As a core market concept developed in the industrial age, market segmentation becomes even more relevant in the information age because of the rich customer information and ubiquitous information technologies. The market segmentation provides business decision makers a very useful perspective to understand the competitive marketplace from the customer's point of view.

Market segmentation is a multicriterion problem but most existing methods solve the problem using single objective optimization methods. By directly tackling the multicriterion nature of market segmentation, this dissertation 1) develops a new conceptual framework, 2) defines a new market segmentation model that unifies existing ones, 3) implements a set of new solution techniques for the unified model, and 4) empirically evaluates the proposed methods using real business data.

1.1 Market Segmentation in the Information Age

Information technologies, especially the World Wide Web and the Internet, have fundamentally changed the marketplace landscape in the past twenty years. Ubiquitous computing devices such as RFID, sensor network and wireless network make the consumer data collection, distribution, and processing much simpler and more economically possible than before. Today's customers are more demanding and more knowledgeable because of the easy access to online product and service information and the information sharing within the customer communities. E-mail, instant message, blog, e-commerce, self-service and online communities are new promotion, distribution and communication channels that allow one business to distinguish itself efficiently from its competitors. As a consequence of the rich customer data, changing customer needs and new marketing tools, academicians and practitioners demand more sophisticated, robust and scalable market segmentation techniques to meet the challenges in market segmentation. After an extensive review of market segmentation literature, Wedel and Kamakura (Wedel and Kamakura 2000) observed that "the development of segmentation theory has been partly contingent on the availability of marketing data and tools to identify segments on the basis of such data. New methodology has often opened new ways of using available data and new ways of thinking about the segmentation problems involved." This dissertation is such an attempt that provides a new conceptual framework and solution techniques to solve problems originating from the multicriterion nature of market segmentation.

Specifically, the solution techniques developed in this dissertation can be seen as an important part of a marketing information system. "A marketing information system consists of people, equipment, and procedures to gather, sort, analyze, evaluate, and distribute needed, timely, and accurate information to marketing decision makers" (Kotler 2003). This dissertation focuses on the information analysis and evaluation with a new conceptual framework. Another distinguished feature of the proposed solution techniques is the synergy of managers' knowledge and data models. Hoch (Hoch 1994) showed that the integration of manager intuition and data model can achieve the optimal result. The segmentation methodology and tools proposed in this dissertation follow this design principle through an interactive and iterative market segmentation process.

1.2 The Role of Market Segmentation

Market segmentation is one fundamental concept of modern marketing (Wind 1978) and is one of the most pervasive activities in both the marketing academic literature and practices (DeSarbo and Grisaffe 1998). There are maybe a dozen popular definitions for what is marketing and the marketing definitions evolve with time, but the original definition of marketing segmentation by Smith (Smith 1956) a half century ago is still valid and widely used today. Wedel and Kamakura's survey (Wedel and Kamakura 2000) on market segmentation revealed more than 1600 references to segmentation in year 2000. The survey (Weinstein 1993) of 203 top-level U.S. marketing executives in 9 industries showed that segmentation strategies were used by two-third of the firms and were found to be significantly more effective than undifferentiated marketing strategy. According to a 2004 management tools survey conducted by Bain & Co. (Rigby 2005), the trend is that executives (960 international respondents) are becoming more customer-centric and market segmentation is ranked 4th in usage and satisfaction among 25 of the most popular management tools and techniques. It is ahead of SCM, TQM, BRP and loyalty management and helps to improve the bottom line results.

Myers (Myers 1996) stated that "one of the most important strategic concepts contributed by the marketing discipline to business firms and other types of organizations is that of market segmentation." Market segmentation plays an important role for business decision makers at both the strategic and the tactical levels. At the strategic level, market segmentation can help answer the fundamental questions of any business including who are our customers, what are their needs, what are the value propositions to meet these needs, and how do we differentiate our products/services in the targeted customer segment(s) from the competitors? Importantly, it can also be used at the tactical level for instance to help make managerial decisions on the four P's (price, product, promotion and place). Wind (Wind 1978) noted that "management must employ the concept of segmentation in all studies, i.e., analyze the data of all studies at the segment level."

Despite the strategic and tactical value of segmentation, its potential has not been fully realized. Recently, McDonald and Dunbar (McDonald and Dunbar 2004) attributed many existing marketing woes to " a lack of understanding of the pivotal importance of market segmentation in successful marketing." Customer Relationship Management (CRM) has emerged as one way to improve the value of segmentation. In fact, after exploring the different definitional aspects of CRM, Payne and Frow (Payne and Frow 2005) defined CRM as "a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments." This suggests that customer segmentation, selection and targeting are the core components of a customer relationship management framework (Winer 2001; Pepper and Rogers 2004).

Though it is possible and in some situations reasonable to segment market in an a priori way, this research considers the more general *post hoc* approach (Green 1977; Wind 1978), i.e., the number and type of segment are data driven and are the results of data analysis. Kotler (Kotler 2003) calls an a priori segment such as young (25-35) middle-income ($50,000 - $80,000) buyers a sector. We also make this distinction in this dissertation.

1.3 The Multicriterion Nature of Market Segmentation

Essentially, market segmentation is a multicriterion (or multiobjective) problem. The initial definition of market segmentation (Smith 1956) said: "Market segmentation consists of viewing a heterogeneous market as a number of smaller homogenous markets in response to differing product preferences among important market segments." It means that the market segments, which should be homogeneous within segments, have to be related to other marketing activity variables, such as response to a marketing mix, to be useful. This view was shared and extended by many marketing researchers (E.Frank, Massey et al. 1972; Myers 1996; Wedel and Kamakura 2000; Kotler 2003) into a set of criteria that determines whether a specific market segmentation is a good one or not. The multicriterion nature of market segmentation can be discussed from three perspectives: the segmentation criteria, the predictive and descriptive models, and the segmentation bases.

1.3.1 Segmentation Criteria

Market researchers often desire both homogeneous market segments (identifiability) and appropriate explanation of response or criterion variables (responsiveness). Identifiability is the extent to which managers can recognize distinct customer groups in the marketplace by using specific segmentation. If segments respond uniquely to marketing efforts, they satisfy responsiveness criterion (Wedel and Kamakura 2000). In addition to the basic identifiability and responsiveness, Frank et al. and Desarbo and Grisaffe (Frank, Massey et al. 1972; DeSarbo and Grisaffe 1998) suggested that the market segmentation should meet the accessibility, feasibility and profitability criteria to implement an effective and efficient market segmentation. Market segmentation meeting those criteria enables a so-called normative segmentation (Claycamp and Massy 1968) that aggregates customers optimally into groups. Generally, identifiability, responsiveness, substantiality, accessibility, stability and actionability are six widely accepted criteria to evaluate whether a segmentation solution is a good one or not (Wedel and Kamakura 2000). The substantiality criterion requires that the size of targeted segments should be big enough to be profitable. Accessibility means that the managers can reach the targeted segment(s) from the available communication or distribution channels. Stability defines that the segments should not change for a time period that allows the execution of a marketing campaign. Actionability means that decision makers are able to formulate an effective and efficient marketing campaign based on the segmentation results. In practice, each criterion is instantiated as a set of optimization objectives and/or constraints.

Additionally, for any particular market segmentation objective, there may be many constraints originating from managerial, institutional, environmental and resource-related restrictions (DeSarbo and Grisaffe 1998). For example, an airline may not want to have more than four customer segments because of the hardware constraints of its flight vehicles.

1.3.2 Predictive and Descriptive Models

Market segmentation methods can be broadly classified into two categories based on whether they use descriptive or predictive statistical methods. Descriptive methods do not distinguish between dependent and independent variables while predictive methods analyze the relationship between a set of independent variables and one or more dependent variables. From the view of segmentation objectives, descriptive methods are optimized for segment identifiability while predictive methods are optimized for segment responsiveness (Frank, Massey et al. 1972). However, a good market segmentation method should meet both criteria (Frank, Massey et al. 1972; Wedel and Kamakura 2000). This is an essential requirement in direct marketing where targeted customers within a segment should be similar to each other with respect to the segmentation bases and should also respond in a similar way to a particular marketing campaign.

1.3.3 Segmentation Bases

Different segmentation bases describe different features of the customer or marketing mix and have different levels of effectiveness regarding the segmentation criteria. For example, geodemographic data has good support for identifiability, substantiality, accessibility and stability but lack actionability and responsiveness. On the other hand, customer benefit data has good support for actionability and responsiveness but only mediocre support for accessibility and stability. In a prediction model, it's common to use one descriptive dimension such as demographic or psychographic variables to identify customers and to predict response variables. In joint market segmentation, more than one segmentation bases is used to take advantage of the benefits of each segmentation basis.

1.4 The Research Questions

Given the multicriterion nature of market segmentation, there are many segmentation methods tackling this problem. This dissertation views market segmentation from both information science and marketing theory points of views. The first research question is how we classify existing market segmentation methods from all these perspectives. The classification helps position this dissertation in a general conceptual framework.

An important piece of this dissertation is the general definition of multicriterion market segmentation using a set of mathematic formulas. The rigorous definition facilitates the development and comparison of different solution techniques. The second research question is what is the unified mathematic model that captures the multicriterion nature of market segmentation? The model should be general enough to unify the existing market segmentation methods. It lays the foundation of the development of segmentation methods and empirical evaluations.

Once the mathematic model is defined, there are two research questions to be answered for it to be useful. At the implementation level, what are the effective and efficient solution techniques for the multicriterion market segmentation problem? At the methodology level, what is the new segmentation methodology derived from the unified mathematic model? The methodology defines the business processes, roles of decision makers, and rules and guidelines of using the system.

Finally and perhaps more importantly, how useful and practical are such a unified model and its implementations? The new model and implementation need to be evaluated empirically in different business settings.

1.5 The Research Methodology

This research crosses the boundaries of management information systems and marketing science. It follows the philosophy of design science (Hevner, March et al. 2004): "The design-science paradigm seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts." We use a problem solving paradigm to find effective solutions to the important marketing segmentation problems.

This research adopts a multi-methodological approach that uses the two design processes (build and evaluate) to produce four design artifacts (construct, model, method, and instantiation) (March and Smith 1995). The methodology is summarized in Figure 1.1.

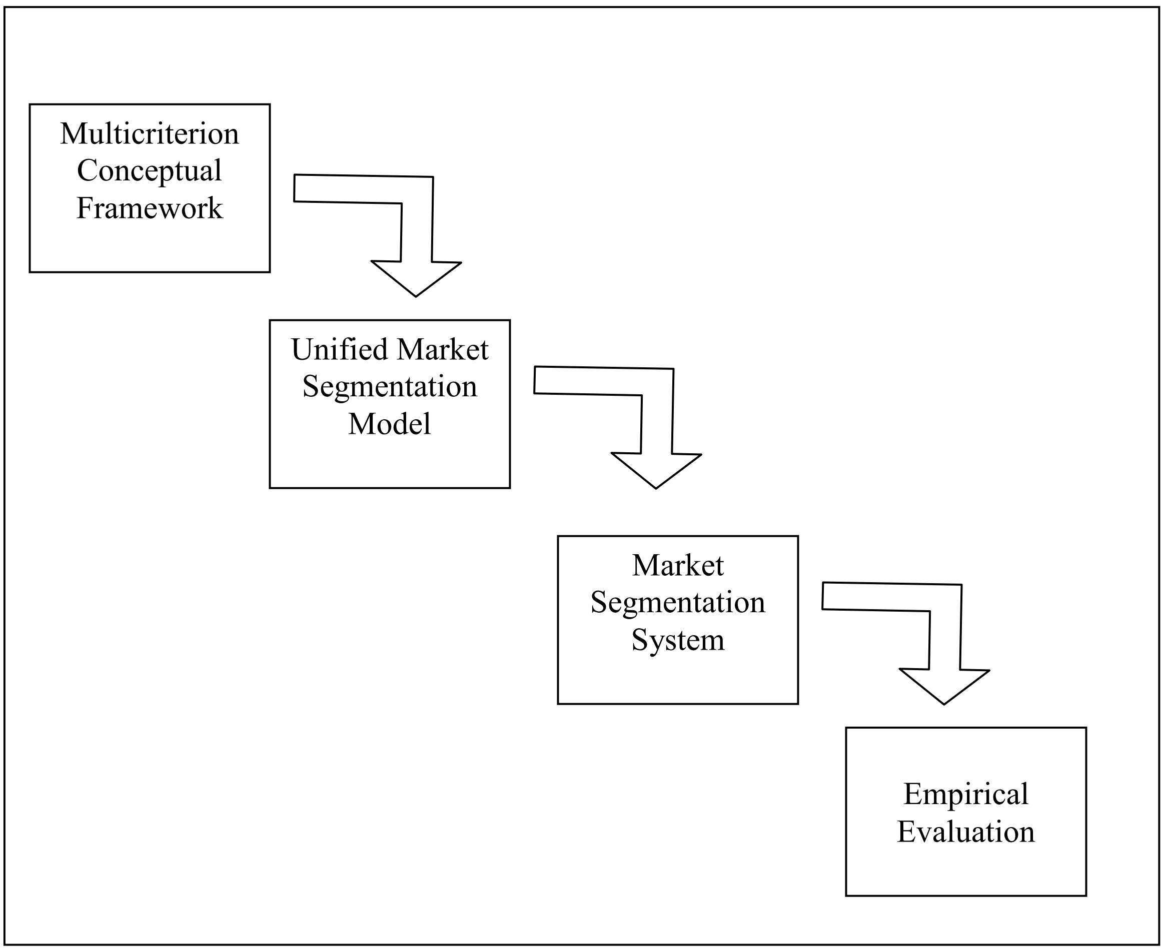


Figure 1.1 Multi-Methodological Approach

More specifically, we build a new conceptual market segmentation framework as well as a rigorous mathematic model for the multicriterion market segmentation problem. A meta-heuristic search algorithm is designed as the method artifact. A fully functional implementation is produced as the instantiation artifact. In the evaluation process, three experiments using real world business data are conducted to evaluate the effectiveness and efficiency of the artifacts. Hevner et al (Hevner, March et al. 2004) established seven guidelines that could be used to evaluate design science research. The evaluation is given in the conclusion section of this dissertation.

1.6 The Dissertation Organization

This dissertation is organized around the four research questions and the corresponding research methodologies. From the information science perspective, Chapter 2 discusses the relationship between clustering and market segmentation and the challenges in finding the “right” number-of-segments value. The computational and practical issues raised in the first half of Chapter 2 establish the foundation for us to understand the existing methods and help to develop the new conceptual framework of those methods in the second half of Chapter 2. A new classification of market segmentation methods is developed from the literature reviews. Chapter 3 starts with a discussion of Pareto optimality and the motivation of using multiobjective optimization. It then defines a rigorous mathematic model for the general multicriterion market segmentation problem.

It shows that a multiobjective-based unified market segmentation framework elegantly addresses the multicriterion nature of market segmentation in all three dimensions (multiple criteria, predictive vs. descriptive, and multiple segmentation bases). Chapter 4 describes the solution techniques of the unified market segmentation model. The implementation consists of multiobjective evolutionary algorithms and automatic solution suggestion algorithms. In Chapters 5, 6, and 7 the unified framework and implementation are evaluated using real customer data sets. The effectiveness and efficiency of the proposed methods are demonstrated in solving real business problems. Conclusions and future research directions are given in Chapter 8.

2 LITERATURE REVIEW

A basic task of market segmentation is to group customers based on similarities in their needs and preferences (Kotler 2003). Clustering is a common tool for this purpose (Punj and Stewart 1983). Clustering could be generally defined as a set of techniques that group entities that are similar in measured characteristics (Jain and Dubes 1988). It aims to maximize the homogeneity within the group while maximizing the heterogeneity among groups. Though necessary, clustering (grouping) is far from enough for any non-trivial market segmentation practice. The intrinsic computational complexity of the clustering and segmentation problem has fundamental effects on any practical solution techniques. Additionally, resource constraints and the computational nature of segmentation are important factors in determining the "right" number-of-segments. The computational complexity implies that any segmentation method is a heuristic one. This chapter first discusses these issues and their implications.

This dissertation investigates the market segmentation problem primarily from an information science point of view. This viewpoint determines the aspects of interest to the author. In a loose sense, we take the combined optimization and data mining approach suggested by Padmanabhan and Tuzhilin (Padmanabhan and Tuzhilin 2003). By approaching the data segmentation from optimization and data mining perspectives, this dissertation investigates existing methods with an emphasis on the assumption about the data, the formulation of the optimization representation, and the optimization solution techniques. After a brief review of existing methods in the single objective optimization context, I discuss the mechanisms that are used to address the issues brought out by the multicriterion nature of market segmentation. The review of these literatures leads us to a new classification of market segmentation methods. It reflects the understanding of how existing methods address the multicriterion nature of market segmentation. As shown in the next chapter, it also enables us to look at the market segmentation problem from a new multiobjective point of view.

2.1 Clustering and Market Segmentation

Market segmentation and clustering are closely related to each other. Besides many other similarities, a fundamental task of market segmentation is assigning customers to different segments (clusters). From the knowledge management point of view, a cluster is a basic knowledge building block for marketers. Each segment is a class of customers that marketers can identify, target, and communicate with. In the early market segmentation research reviewed by Wind (Wind 1978), clustering was considered to be almost synonymous with market segmentation. The reviews by Punj and Stewart (Punj and Stewart 1983) as well as Arabie and Hubert (Arabie and Hubert 1994) had comprehensive discussions on the role of cluster analysis in market segmentation.

However, we believe one can **NOT** equate market segmentation to clustering. First, at the very basic business application level, decision makers do not segment customers for "clustering" purposes only. They want actionable segments that will let them formulate an effective marketing campaign in an objective way. As discussed in the previous chapter, market segmentation by its definition is related to the marketing mix. When launching a market promotion or advertisement, the customer response of each segment is as important as the segment profile itself, if not more. Many segmentation methods such as clusterwise regression methods put more weight on the regression than on the clustering. One might argue that no response data is available in the segmentation research for a new product. Nonetheless, decision makers can segment customers based on more than one segmentation basis (such as psychographics, demographics, and old product usage patterns) using a joint segmentation (Morwitz and Schmittlein 1992) method which could give more insights for new product design/development. Once the product prototype is ready, an immediate pilot user study would allow predictive segmentation to guide customer selection and product positioning/targeting. In both predictive and descriptive segmentation, clustering is only part of the problem. Second, at the general conceptual level, clustering only addresses the identification criterion (Brusco, Cradit et al. 2003) while other criteria (Wind 1978; Wedel and Kamakura 2000) have to be addressed by augmented methods.

As a result, methods that only consider the clustering goal of market segmentation provide a partial solution to the segmentation problem at best. As discussed in the next chapter, multistage or transformation methods are often used in practice to solve the complete puzzle of market segmentation.

2.2 The Computational Issues of Clustering

Though it is only part of the segmentation problem, clustering by itself is vaguely defined. The clustering process is hard and fuzzy (Jain, Murthy et al. 1999). The vagueness lies in two aspects: the measurement of so-called “homogeneity” or “similarity” and the clustering process.

Problems of determining the appropriate similarity measure in marketing were discussed in several studies (Green, Frank et al. 1967; Frank and Green 1968; Punj and Stewart 1983). From a number of empirical experiments, they found that each similarity measure has different characteristics and different distance measures lead to different clustering results. Skinne (Skinne 1978) identified the three aspects of similarity measures: elevation, scatter, and shape. In a rough sense, elevation could be thought of as the mean of all attributes. Scatter is about deviation, while shape is about the direction (up/down) of the data. The most important finding of (Skinne 1978) is that a distance or similarity measure may cover one or more aspects. For example, Euclidian distance only accounts for elevation and scatter and a correlation coefficient accounts only for scatter. As a result, different clustering techniques based on different similarity measures reflect one or the other aspect of the problem. For example, the hierarchical clustering method works well if the similarity measure is in the form of proximity matrices.

Even the probabilistic clustering methods make an assumption of similarity measure, though in an indirect way. Those methods assume that the overall distribution of the data is a mixture of probability distributions, each being a different cluster (Fraley and Raftery 1998). Common data assumptions are multinomial distribution for categorical scale or multivariate Gaussian for interval or ratio data. Each assumption has the intrinsic similarity aspects described above. For example, Gaussian distribution measures elevation (mean) and scatter (variance). The probabilistic clustering methods are powerful if the data assumptions are right.

The interesting thing is that a similar but more restricted impossibility theory of clustering process has been proved by Kleinberg (Kleinberg 2002) recently. It is intuitive to think of three desired properties of any clustering process. Scale-invariance property means that changing the unit of distance measure should not change the clustering result. Richness requires that a clustering process should be able to generate all possible partitions of clustering entities. Finally, consistency is satisfied when the clustering result stays the same when we increase the distance among clusters and decrease distances within clusters. He showed that there is no clustering process that can satisfy all three properties simultaneously.

The three aspects of similarity measure and impossibility theory of clustering process suggest that any similarity measure and clustering process carries special characteristics. For any given clustering problem, the problem assumptions, similarity measures and solution techniques should match. It is not a surprise that there is no generally accepted criterion for determining the quality of clustering methods. Different methods have different advantages and disadvantages. Complementary measures and processes often give more comprehensive understanding of the problem than a single measure and/or single process. User knowledge of the problem domain is very important for an unsupervised learning task such as clustering.

2.3 The Computational Complexity of Market Segmentation

The rich customer data set and short market response time focus attentions on the computational complexity of market segmentation. The clustering problem in its basic partition format requires a permutation of all entities. Kleinberg (Kleinberg, Papadimitriou et al. 1998) defined a general segmentation problem as partitioning entities into a number of segments. The goal is to maximize the sum of functions defined on those segments. He proved that even the most trivial optimization problems become NP-complete if they are defined in a segmentation form.

Krieger and Green (Krieger and Green 1996) defined the market segmentation problem as a 0-1 programming problem. The computational complexity of the problem is NP-hard.

Consequently, clustering, market segmentation or general segmentation problems can not be solved in polynomial time in their general forms. Existing methods either transform the problem into an easy to solve version or apply heuristic techniques to solve the problem.

2.4 Determining the Number-of-Segments

Wedel and Kamakura (Wedel and Kamakura 2000) observed that "the problem of identifying the number-of-segments is still without a satisfactory statistical solution." This issue can be examined from two perspectives: the managerial perspective and the theoretical perspective.

From a managerial perspective, the "number-of-segments" is usually discussed in the context of macro, micro, and one-to-one market segmentation (Rubin 1997; Payne and

Frow 2005). Some companies are moving into the one-to-one segmentation arena (Peppers, Rogers et al. 1999). These can be seen as extreme cases where there are enough managerial and institutional resources allowing perfect customer discrimination. In some situations, the number-of-segments can be a very simple decision. For example, many software companies only support two or three product editions in one product line because each new product edition requires significant development, marketing, and support effort. To some extent, the choice of the number-of-segments can be purely judgmental in practice. As a rule of thumb, McDonald and Dunbar (McDonald and Dunbar 2004) found that 5 to 10 segments are common while Myers (Myers 1996) suggested 3 to 10 segments. Nonetheless, decision makers always want more objective evidence from data.

Theoretically, recent research on the clustering problem (a sub problem of market segmentation) has examined the problems in determining the optimum number-of-clusters. Though model-based clustering (Fraley and Raftery 1998; Tibshirani, Walther et al. 2000) and mixture segmentation models (Wedel and Kamakura 2000) use statistical inferences to determine the number of clusters, the work of Kleinberg et al. (Kleinberg, Papadimitriou et al. 1998; Kleinberg 2002) stated that 1) segmentation is a NP-complete problem and 2) there is no single clustering function satisfying the scale-invariance, richness, and consistency requirements of clustering. Consequently, Handl and Knowles (Handl and Knowles 2004) used two complementary clustering functions (compactness and connectedness of clusters) to detect the number of clusters. Alternatively, one can examine the issue from a simple cluster measurement viewpoint. A fundamental issue in clustering is that the measurement of similarity is multi-dimensional: elevation, scatter and shape represent different aspects of the data attributes and usually is not covered by a simple measure (Skinner 1978). These findings also explain the problems in cluster validation.

Market segmentation requires more than clustering. In predictive segmentation, the criterion of the predictive power is as important as the criterion of the segment homogeneity. Generally speaking, as the number-of-segments increases, the within-segment homogeneity usually increases but the predictive power may increase or decrease independent of the within-segment homogeneity. Joint segmentation consists of clustering on multiple segmentation bases, in that, each can be thought of as an independent clustering problem and an overall tradeoff has to be made in selecting the "right" number-of-segments. The multicriterion nature of market segmentation means that determining the "right" number-of-segments is a multicriterion decision and often involves marketers' domain knowledge.

Consequently, from both managerial and theoretical perspectives, decision makers would like to see a set of segmentation solutions that have different numbers-of-segments and select the "right" one based on multiple, often conflicting, objectives. Managerial and resource constraints play an important role in determining the most suitable or "best" number-of-segments. A good segmentation method should provide decision makers with a holistic segmentation view that enables them to find the best solution for a specific business setting.

2.5 A Brief Historical Review

The early stage market segmentation research is characterized by a simple descriptive segmentation model as well as the lack of customer data and computation power. Though the market segmentation concept was widely accepted and was used in a wide range of business scenarios since the 1950s, before 1970 clustering-based segmentation models prevailed in marketing literature (Green, Frank et al. 1967; Wind 1978; Punj and Stewart 1983). This is not a surprise given the close relationship between clustering and segmentation. The research goal then focused on understanding customers from both general customer characteristics and situation specific customer characteristics (Frank, Massey et al. 1972). General customer characteristics consist of general descriptions regarding the customer as a motivated human being. The segmentation bases include but are not limited to demographic, socioeconomic, personality and life style, media usage, and benefit sought variables. The situation customer characteristics are associated with the 4Ps (product, price, promotion and place) of marketing and include purchase and loyalty patterns, price sensitivity, perception and preferences, and others. After 1970, segmentation methods focused more on the relationships between a dependent variable and a set of predictors. Automatic interaction detection (AID), chi-squared automatic interaction detector (CHAID), classification and regression trees (CART), and clusterwise regression became popular in response to the shifted focus on predictive models. Recently, complex probabilistic model-based methods and heuristic search-based methods are common tools for marketing researchers. The state of the art of probabilistic methods is the finite mixture model (FMM) and its variants. Artificial neural network (ANN), simulated annealing, and genetic algorithms are very useful to deal with the computation complexity issues in segmentation.

2.6 Descriptive and Predictive Methods

Due to its theoretic importance, the descriptive and predictive classification of market segmentation methods provides a unique view for looking at the multiobjective nature of market segmentation. Simultaneously clustering and predicting is a long standing problem facing marketing researchers.

Commonly used descriptive methods are clustering-based methods such as variations of K-means, hierarchical clustering techniques (Punj and Stewart 1983; Jain and Dubes 1988; Arabie and Hubert 1994) and Self-Organizing Map (SOM) (Kiang and Kumar 2001; Kiang, Hu et al. 2006) methods. If assumption of the density distribution can be made for the data set, probabilistic models are more appropriate for clustering tasks with the additional advantages of statistical inference (Chesseman 1996; Fraley and Raftery 1998). Artificial neural network (ANN), evolutionary approaches, and other search-based approaches recently become common tools because of the large data set and increased computation power. A good review of clustering methods can be found in (Jain, Murthy et al. 1999). The clustering-based methods focus on customer within segment homogeneity and usually ignore the relationships between segmentation bases and exogenous response variable(s). Joint segmentation (Ramaswamy, Chatterjee et al. 1996) tries to segment customers on more than one basis or dimensions. Two or more dimensions are of different types, e.g., demographic descriptors and benefit descriptors. Finite mixture models can be use to solve the joint segmentation model by maximizing the joint probability of multiple predictor dimensions. Andrews et al (Andrews, Ansari et al. 2002) compared Hierarchical Bayes with the finite mixture model in conjoint analysis applications. Descriptive models generate good solutions in terms of within segment customer homogeneity.

A variety of response modeling techniques such as clusterwise regression model (Spath 1979), clusterwise logistic model, CHAID, and mixture regression model (Wedel and DeSarbo 1994) have been widely used as predictive methods. Wedel and Kamakura (Wedel and Kamakura 2000) gave a comprehensive review of the state of the art of mixture models. Predictive methods usually result in a better predictive model for each segment than for the whole population. But the within segment homogeneity of independent variables is somewhat low.

Simply speaking, the single objective-based market segmentation methods achieve a good result in one criterion at the expense of the other one. Many traditional descriptive clustering methods and predictive clusterwise methods have been enhanced to overcome their intrinsic shortcomings. Vriens et al (Vriens, Wedel et al. 1996) provided an extensive conceptual review and empirical comparisons of modifications of clusterwise regression methods and latent class models. Those methods use the two types of approach discussed below and share the same problems with them.

2.7 Two Approaches for Multiobjective Market Segmentation

An intrinsic problem associated with the multicriterion nature of market segmentation is that multiple criteria and their corresponding optimization objectives are often antagonistic. For example, a segmentation method optimizing the descriptive homogeneity usually leads to a less satisfied predictive segmentation solution, or vice versa. The multicriterion nature of market segmentation raises many issues that cannot be addressed appropriately by traditional market segmentation methods such as K-means and clusterwise regression because they only optimize one objective. As a result many methods have been proposed to address the multicriterion requirement of market segmentation (Krieger and Green 1996; DeSarbo and Grisaffe 1998; Wedel and Kamakura 2000; Brusco, Cradit et al. 2002; Brusco, Cradit et al. 2003). In general there are two approaches for multicriterion customer segmentation: multistage approach and transformation approach. In the multistage approach, the multiobjective market segmentation problem is treated as multiple single criterion problems and each criterion is addressed in a stage. Therefore multiple criteria are addressed in a stage-by-stage approach. Alternatively, in a transform approach, the multiobjective problem is transformed into a single criterion problem by a scalar function such as a weighted-sum function. Then the problem is solved in a single criterion way. An early study by Vriens et al. (Vriens, Wedel et al. 1996) reviewed and compared several traditional multistage and transformation methods in the context of conjoint segmentation. The following review focuses on conceptual level issues and recent advances.

2.7.1 Multi-stage Approach

The multi-stage approach allows researchers to deal with one criterion at one time. There are many mature single criterion optimization methods that can be used in each stage. Green and Krieger (Green and Krieger 1991) used a two stage approach to relate market segmentation with a firm's product/marketing mix. In the first phase, the researcher describes the among group heterogeneity in terms of product attribute preferences. In the second phase, the implication of preference heterogeneity is evaluated and the result is used to optimize the product portfolio. Krieger and Green (Krieger and Green 1996) used another two-step approach. In the first stage, the K-means method is used to cluster respondents into groups that are optimized for group identifiability. A second phase of heuristic algorithm is used to improve the responsiveness of segments with a limitation on the decrease rate of within group homogeneity. The disadvantage of the multi-stage approach is that the information found in one stage is not used by the other stages because of the separated processing phases. The algorithm is not efficient in the sense of information sharing. The order of objective optimization often matters. If one changes the order of optimization objective, the result usually is significantly different. Further more, because each stage optimizes a single objective, the final result is often suboptimal with regard to all objectives.

2.7.2 Transformation Approach

The other widely used approach is to transform the multiobjective problem into a single objective problem. Consequently the problem can be solved by many well-studied single objective optimization methods. Based on the multicriterion and multiple constraints in the segmentation problem, DeSarbo and Grisaffe (DeSarbo and Grisaffe 1998) defined a total utility function to integrate the different criteria and applied single objective combinatorial optimization approaches. Most mixture models described in Wedel and Kamakura (Wedel and Kamakura 2000) also fall into the transformation category because multiple criteria are integrated into a single maximum likelihood function for the joint probabilistic distribution. The single objective Newton-Raphson or expectation-maximization (EM) heuristics are used to estimate the segment membership and other mixture distribution parameters. Recently, Brusco et al (Brusco, Cradit et al. 2002) presented a simulated annealing heuristic (SAH) algorithm for solving the bicriterion partition problem in market segmentation. The bicriterion method was extended in (Brusco, Cradit et al. 2003) to handle the three criteria segmentation problems using the weighted sum function with three terms. In both methods, a weighted-sum function is used to convert the multicriterion problem into a single criterion problem. The SAH algorithm is executed to find the optimal solution for the scalar goal as a result of an optimal combination of weighting parameters. If the result has an unacceptable result for any of the two objectives, the weights of the objectives are changed and SAH runs iteratively. Gradually an appropriate compromise is achieved for both objectives. Though this method simultaneously evaluated both objectives and simulated annealing avoids a local optimal solution, there are many issues to be solved. First, the trial-and-error approach to set weight value for different criterion is far from efficient. Second, the step length of weight value and the lower bound of objectives are subjective settings and might miss a good solution. Additionally, because of the trial-and-error approach, it is not scalable for the multiple objectives problem because the combination of weights increases exponentially with the number of objectives.

Nonetheless, the transformation approach has many problems to be resolved. First, it is difficult, if not impossible, to define an appropriate total utility or weighted sum function. How do we set the weight of each objective? What if objectives are incommensurate? For example, one criterion is the within-segment homogeneity measured by within-segment sum of variance and another criterion is the predictive power in logistic regression measured by maximum likelihood. Additionally, the transformation procedure may put unnecessary limitations on the search space; thus the global optimal solution could be missed “by transformation.” Freitas (Freitas 2004) gave an example in which there are three two-dimensional data: (9.5, 5.0), (7.0, 7.0), and (5.0, 9.5). Suppose larger is better, then it is impossible to find a linear combination of weights that selects the (7.0, 7.0). The data could be a favorite solution in practice because of its good balance in both measures.

2.8 Problems of Existing Methods

In general, current market segment methods based on either multi-stage or transformation approaches have several limitations that often make market segmentation results far from adequate for both academic researchers and market practitioners. The multistage approach is not efficient and effective in segmenting customers because it considers each objective individually while the multiple objectives are highly related to each other by sharing the common customer segment membership information. For the transformation approach, it is often difficult, if not impossible, to find a total utility function to represent the tradeoffs of multiple objectives. Usually many subjective decisions have to be made up front to define the transformation function that transforms multiple objectives into a single optimization objective. To make things worse, such methods often put unnecessary constraints on the searching space that often result in suboptimal solutions. Furthermore, instead of a single optimal solution, the multiobjective market segmentation problem has many, theoretically an infinite number of, solutions that meet the constraints and represent different tradeoffs of the multiple objectives. Given the huge search space and many possible solutions, users of the algorithm have to run the algorithm many times with many subjective settings to find one or more appropriate results. The process is time consuming and error prone. Finally, many of the existing market segmentation methods such as K-means clustering, heuristic clusterwise regression and the EM algorithm stop at local optimal solution(s).

Simply, the drawbacks of existing market segmentation methods could be attributed to three problems. First, these methods use a single objective approach to address the intrinsically multiobjective problems. Second, usually one suboptimal solution is generated by the single objective optimization methods. Finally, many existing methods use methods that generate local optimal solution(s).

2.9 Discriminative vs. Generative Methods

In addition to the mechanism for solving multiobjective problems, assumptions about the segmentation data play an important role in segmentation methods. Similar to the clustering method classification scheme proposed by Zhong (Zhong and Ghosh 2003), market segmentation methods can be classified into discriminative (or distance/similarity-based) approaches and generative (or model-based) approaches from their computational assumption about the nature of data. Discriminative methods calculate distances or similarity between customers and segment customers based on these measures. K-means, hierarchical clustering, and Self-Organizing Map are typical discriminative clustering methods. Generative methods assume customers are from different statistical models and try to find the parameters of the corresponding models. Recently, the generative mixture model has gained in popularity among market segmentation researchers because it enables statistical inferences (Wedel and Kamakura 2000).

Each type has its advantages and disadvantages. Because of the direct optimization of within-segment customer similarity, discriminative-based segmentation methods are usually efficient and intuitive. However, the results usually are used as-is and no statistical inference could be drawn from the results. There are several advantages of generative methods. If the distribution assumption of data is correct, they usually generate better results than discriminative methods. The results are more interpretable and enable statistic inference. But generative methods such as finite mixture model are computationally expensive when the number-of-segments is big or there are many segmentation variables (Wedel and Kamakura 2000). The scalability becomes more and more important as customer data collection becomes easy and cheap these days.

Both the discriminative and generative approaches may be formulated as a data mining and/or optimization problem (Padmanabhan and Tuzhilin 2003) whose solutions need to have confidence and support, information content and unexpectedness (Kleinberg, Papadimitriou et al. 1998). A discriminative heuristic optimization method usually has better scalability and simplicity than a model-based approach if the goal is to find a set of solutions for the multicriterion segmentation problem.

2.10 Classification of Market Segmentation Methods

A unique feature of this dissertation is the integrated view from both marketing research and information systems. For the concerns of this research, the classification of market segmentation methods has three dimensions that are depicted in Figure 2.1. The classification is based on relationships among segmentation bases, assumptions about the data, and implementation mechanisms for multiobjective optimization.

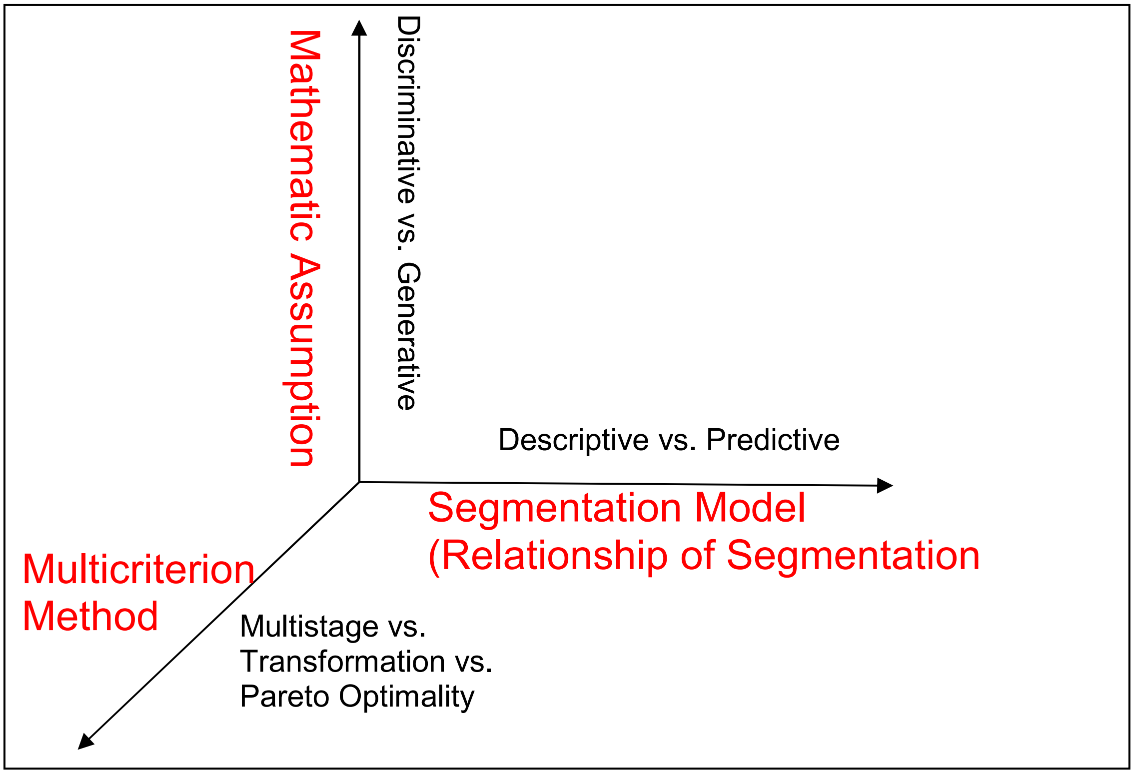


Figure 2.1 Classification of Market Segmentation Methods

The first dimension is the traditional view of market segmentation that is based on the relationship of segmentation bases. This dimension classifies segmentation methods into descriptive and predictive categories depending whether or not they distinguish between independent and dependent variables. In other word, it is about the relationships among segmentation bases. Descriptive methods do not distinguish between dependent and independent variables. It is worth pointing out that more than one segmentation basis is possible in the descriptive model. It is called joint segmentation (Ramaswamy, Chatterjee et al. 1996) in marketing literature. Predictive methods analyze the relationship between a set of independent variables and a set of dependent variables. Each type has its own characteristics and challenges. Descriptive models deal with the fundamental problem of assigning customers into segments that are identifiable easily from the segment profile. Identifiability and accessibility are basic requirements for the descriptive models. Discriminative and generative clustering methods are common tools for descriptive segmentation. Predictive models emphasize the relation homogeneity of response variables and predictors. The result is more actionable because it enables segment-based prediction of marketing response. The predictive model optimizes the responsiveness criterion of segmentation. The predictive power can be measured by common goodness-of-fit measure in a typical regression model. Alternatively, as shown in (Krieger and Green 1996; Brusco, Cradit et al. 2002), the predictive power can be measured by homogeneity of dependent variables if it is appropriate for the specific application.

The second dimension is based on the assumptions of the segmentation data. In generative (model-based) methods, the data is assumed to be a mixture of data with different statistical distribution parameters. In discriminative methods, no such assumption is made and the similarity or distance of the data is directly calculated and used for segmentation. The pros and cons of each assumption are discussed in the previous section.

Tackling the multicriterion nature of segmentation problems defines the third dimension. This dimension can be thought of as the implementation strategy of the solution techniques and is independent of the other two dimensions. As discussed in the previous sections, most existing market segmentation methods can be classified as either multi-stage or transformation approaches. Without loss of generality, a single objective method is classified as a transformation approach in the sense that other segmentation criteria are weighted as zero. The Pareto optimality type refers to a class of solution techniques that optimize multiple objectives simultaneously based on the notation of Pareto optimality. This new type of methods is the theme of this dissertation and will be developed and discussed in following chapters.

Because of the orthogonal nature of these three dimensions, it's easy to categorize a market segmentation method though some hybrid cases are common because of the multicriterion nature of segmentation. For example, a simple K-means or hierarchical clustering method is a descriptive, discriminative and transformational method. Clusterwise regression model (Spath 1979) is a predictive, generative, and transformational approach. Multi-stage or transformation methods may take multiple values in each dimension. The simulated annealing heuristic (SAH) segmentation (Brusco, Cradit et al. 2002) is a combination of predictive, descriptive, discriminative and transformation because it combines predictive and descriptive objectives into a weighted sum function. Mixture models described in Wedel and Kamakura (Wedel and Kamakura 2000) are generative methods. Practically, most mixture models are predictive and transformational methods but other combinations are possible. For example, a mixture model can be applied in joint segmentation (Morwitz and Schmittlein 1992) that is a descriptive method.

The classification is important because it gives hints for the possible solution techniques. Because most existing segmentation methods are either multi-stage or transformational methods, it is helpful to develop and evaluate Pareto optimality-based solution techniques. Additionally, those solution techniques can be either discriminative or generative and can be used in both predictive and descriptive segmentation models. The three dimensions are independent of each other and can be assembled in any combination – the multicriterion nature of market segmentation calls for a unified market segmentation model.

3 A UNIFIED MARKET SEGMENTATION MODEL

The multicriterion nature of the market segmentation problem and the limitations of the existing approaches call for a new segmentation model and new solution techniques. The problems identified in the previous chapter, to a great extend, can be attributed to a root problem of "lost in adaptation" -- adapting multiobjective problems to existing single objective solution techniques not only raises many problems but also loses many benefits by solving the problem directly. The multi-stage approach is inefficient and ineffective because multiple objectives are not optimized simultaneously. Transforming multiple objectives into one is difficult because it is hard to set weight for each objective without seeing the final solutions. Even if multiple objectives are commensurate and their weights are clear in the business domain, the transformation approach may lose the global optimal solution by its definition. Maybe one of the biggest drawbacks in most existing methods is that the single objective method only gives a single solution even though the multiobjective problem has many feasible solutions.

Therefore instead of adapting the multiobjective problems to existing single objective optimization methods, this dissertation develops new multiobjective optimization methods that are dedicated to the multicriterion market segmentation problems. Specifically, what is needed is a multiobjective market segmentation model and new segmentation solution techniques that efficiently and effectively tackle the multiobjective nature of market segmentation problems. As a result of this multiobjective optimization definition, the new approach generates a set of possible solutions that represent different tradeoffs among multiple objectives. This section describes the multiobjective market segmentation model. As shown in the following sections, the multiobjective model elegantly integrates the descriptive and predictive market segmentation models. Additionally, it incorporates multiple segmentation criteria.

3.1 Multiobjective Optimization and Pareto Optimum

To avoid the drawbacks of multi-stage and transformation methods, we propose to tackle the multicriterion nature of market segmentation directly from a multiobjective optimization point of view. In general, the multiobjective optimization problem can be defined (Osyczka 1985) as the problem of finding:

> *A vector of decision variables which satisfies constraints and optimize a vector function whose elements represent the objective functions. These functions form a mathematic description of performance criteria which are usually in conflict with each other. Hence, the term "optimizes" means finding such a solution which would give the values of all the objective functions acceptable to the decision maker.*

Unlike the single optimum approaches, the term "optimize" with multiobjective is defined as finding acceptable solutions regarding the multiple objectives. It is worth pointing out that for a NP-hard multiobjective optimization problem like market segmentation, it is practically impossible to find the optimal solution(s) for a non trivial problem. Depending on the nature of problem, even determining whether a given solution is optimal or not could be a problem without a deterministic answer. In market segmentation, weighting over a number of conflicting objectives (either descriptive and predictive objectives or multiple objectives derived from multiple criteria) is often difficult and subjective. Essentially, for antagonistic multiple objectives, there is more than one "optimal" solution. In a set of feasible solutions, one solution is optimal in certain objectives but it is at the expense of being less optimal in one or more objectives. The solutions are acceptable because they represent tradeoffs among multiple conflicting objectives.

Formally we use the concept of "Pareto optimality" to define the acceptable solutions (Pareto 1896). This definition says that a solution is Pareto optimal if there exists no feasible vector of decision variables which would improve some criteria without simultaneously making at least one other criterion worse. As a result, this concept often finds a set of solutions called Pareto optimal set for a multiobjective optimization problem. Each solution in the Pareto optimal set represents a specific tradeoff of multiple objectives.

Suppose that we want to minimize the within segment heterogeneity for both identifiability and responsiveness criteria in a segmentation problem. Because the two measures are usually antagonistic, one specific segmenting of customers can not be optimized for both objectives. A possible solution set is shown in Figure 3.1.

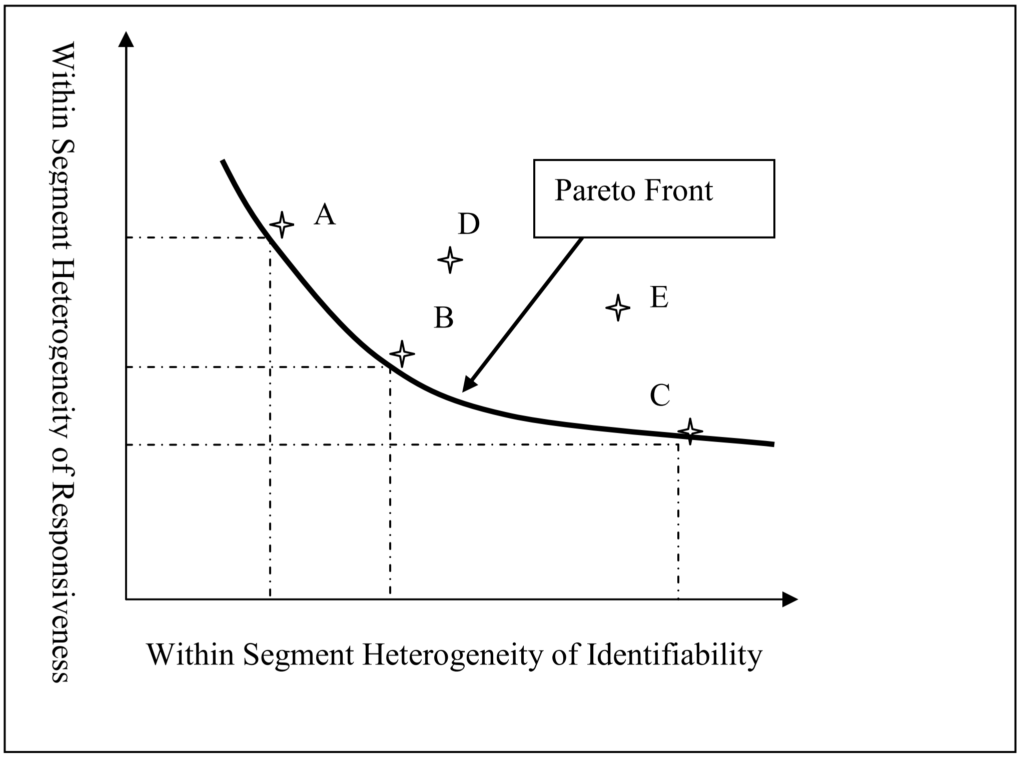


Figure 3.1 Pareto Optimality and Pareto Front

In Figure 3.1, solutions A, B and C are Pareto optimal because there is no such solution whose values are better than them in both dimensions. In other words, A, B and C are not dominated by any other points. Solution A, when compared with B and C, is better in identifiability but worse in responsiveness. However, solutions D and E are not Pareto optimal because solution B is better than them in both measures. As a result, we say D and E are dominated by other points and are not Pareto optimal. All non-dominated solutions are Pareto optimal and the set forms a surface (two dimensions) that is called the Pareto Front. Practically, the goal of a multiobjective optimization algorithm is to find a representative subset of the Pareto optimal set.

The concepts of multiobjective optimization and Pareto optimal solution set bring a new perspective to multicriterion market segmentation. What we are looking for is to find a set of (Pareto optimal) solutions that represent different tradeoffs among multiple conflicting objectives to the decision maker. There are several benefits of this perspective. First, Pareto optimal avoids the difficulties of weighting or combining one criterion with another, sometimes incommensurate, criterion. This is a severe problem of the transformation approach. Second, there is no need to make the up-front tradeoff in some parameters such as objective weight to do the search. Pareto optimal problems can be defined directly and independently in terms of multiple objectives and decision variables. Finally but no less importantly, the goal is to find a set of Pareto optimal solutions. This property, with the aid of visualization tools, significantly improves the productivity of the decision making process.

3.2 A Unified Market Segmentation Model

The multicriterion market segmentation problem can be defined as a multiobjective optimization problem with a set of constraints. Without loss of generality, assume the problem is to minimize a set of objectives.

C = the set of segments in a segmentation solution, indexed by $c = 1,...,K$, K is the number of segments in the solution; different solution may have a different K value. K or a range of K is specified as a constraint.

z_i = the segment membership of customer i; $i = 1,...,I$, I is the number of customers.

$z = [z_1, z_2,..., z_I]$ is a vector of decision variables that represent a segmentation solution for all I customers;

(1) $F(z) = [f_1(z), f_2(z),..., f_M(z)]$ is the objective vector to be optimized; M is the number of objective functions with $M \geq 1$;

(2) $G(Z) = [g_1(z), g_2(z),..., g_P(z)]$ is the constraint vector to be satisfied; P is the number of constraint functions with $P \geq 1$; P is bigger than or equal to 1 because at least the number of segments has to be specified.

(3) A specific segmentation solution $z^* = [z_1^*, z_2^*,..., z_I^*]$ is Pareto optimal if there does not exist another solution z^l such that

i. z^l satisfies (2) and

ii. $f_j(z^l) \leq f_j(z^*)$ for all $j = 1,2,...,M$ and

iii. $f_k(z^l) < f_k(z^*)$ for at least one k

(4) The goal is to find one (or more) segmentation solution(s) $z^* = [z_1^*, z_2^*,..., z_I^*]$ that is (are) Pareto optimal with regard to the objective vector (1) and satisfies constraint vector (2).

For a NP-hard problem such as market segmentation, usually it is impossible to know whether the generated solution is a real Pareto optimal solution or not. For a given multiobjective optimization solution technique, the generated solutions are the best it can find. Though we validate their quality using alternative methods and call them Pareto optimal, in most scenarios they still represent an approximation we could achieve.

The above model definition is very general in the sense that it abstracts very basic elements of any segmentation problems. The decision variables are the segment membership of each customer and the number-of-segments (if a range of number-of-segments is specified as a constraint). The goal of finding a set of Pareto optimal solutions reflects the reality that practically there are many acceptable solutions for a business problem. Additionally, the set of solutions offers another benefit that is missing in the single objective optimization method: when defining a specific instantiation of the above market segmentation problem, the decision makers have the flexibility to either segment with less constraints or segment with as many constraints as possible. The former approach gives the decision makers more options while the later one works better when there are many hard constraints.

The model is general and reflects the unified view of market segmentation in a number of aspects:

- By this definition, the single criterion market segmentation problem is a special case where there is only one objective function and often a single optimal solution exists.

For example, clusterwise regression (Spath 1979) could be represented as $M = 1, P = 1$ and $f_1(X, Y, z) = \sum_{k=1}^{K} ESS(k)$, where $ESS(k) = (Y^k - X^k b^k)^T (Y^k - X^k b^k)$. Y^k and X^k are the response vector and descriptive matrix of segment k correspondingly. b^k denotes the regression parameter vector. K-means clustering optimizes the within segment variance. A variety of finite mixture models optimize maximum likelihood. All can be represented by $M = 1, P = 1$ and their corresponding objective functions. The only constraint is the number-of-segments.

- The segment membership could be non-overlapping, overlapping (Carroll 1976) and fuzzy (Hruschka 1986). The actual value of customer membership (z_i) is problem specific. For example, if z_i is a single segment identifier defined on the domain of C, the segmentation solution is a partition of all customers. If z_i is a set of segment identifiers, the solution is an overlapped segmentation. If z_i is a vector of membership probability for each segment, the segmentation is a fuzzy one.
- The number-of-segments K is an attribute of a specific segmentation. A specific K or a range of K must be specified as a constraint for a segmentation problem if the problem domain enables decision makers to do so. Managerial, institutional, or resource constraints are good sources for determining the value of K.
- The result is not a single solution but a set of Pareto optimal solutions. The solution set may consist of segmentations with different number-of-segments. This reflects the multicriterion nature of the market segmentation problem.

- The constraint can be defined on decision variables and on objective functions. For example, a constraint could be that the minimum segment size is 50 so that the substantiality criterion is met. A constraint on objective function could be that the minimum R squared is 0.2 for any segment level linear regression.

In a descriptive segmentation model, objective function is often defined by one within segment homogeneity measure such as the ratio of the sum of within segment variance to the total sum of variance. In a predictive segmentation model, the objective function is often defined by a predictive accuracy measure such as R squared. Therefore by allowing multiple optimization objectives, this new market segmentation elegantly unified the predictive and descriptive segmentation model.

The general segmentation definition in Desarbo and Grisaffe (DeSarbo and Grisaffe 1998) is close to our formulation but has fundamental differences. Desarbo and Grisaffe (DeSarbo and Grisaffe 1998) discussed the multicriterion and multiple constraint nature of market segmentation. They pointed out the existence of a set of acceptable Pareto optimal solutions. Though both definitions allow multiple optimization objectives and multiple constraints, there are two fundamental differences between them. First, the market segmentation defined here is an explicit multiobjective optimization problem. A single optimization function, though desired for algorithm simplicity, is usually hard to build and is not necessary. Nonetheless, given that there are many good single objective combinatorial optimization algorithms, Desarbo and Grisaffe (DeSarbo and Grisaffe

1998) defined a general total utility function that integrates the multiple objectives. Second, the multiobjective definition results in a set of Pareto optimal solutions that represent a variety of tradeoffs among the antagonistic optimization objectives. This feature shifts the burden of upfront compromise to a post optimization stage in which the decision makers already have a set of Pareto optimal solutions to play with. For each solution, the member information, segment profile and solution objective values are known. Consequently, the most important tradeoff information among solutions is made explicitly at this stage. This gives the much desired insights and flexibility to the decision makers.

Other multicriterion market studies such as (Krieger and Green 1996; Brusco, Cradit et al. 2002; Brusco, Cradit et al. 2003) only defined problem specific objectives that were either addressed by a multistage approach (Krieger and Green 1996) or were integrated into a weighted sum function (Brusco, Cradit et al. 2002; Brusco, Cradit et al. 2003). They represented the recent effort that explicitly acknowledged the multicriterion nature of market segmentation and developed new methods to solve the issues. They did not define the problem based on the notion of Pareto optimality and a set of acceptable solutions. Consequently, they have the drawbacks discussed in the previous chapter.

3.3 The Market Segmentation Methodology

There are many segmentation methodologies designed to guide the whole lifecycle of segmentation studies. Among others, Wind (Wind 1978) defined a five phase segmentation methodology that consists of problem definition, research design, data collection, data analysis, and data interpretation and implementation of results. The five phases represent the basic tasks performed by marketing researchers and practitioners. Myers (Myers 1996) proposed a similar six step procedure with a focus on data analysis: decide on segmentation bases, choose data analysis methodology, identify segments, profile segments, target segments, and develop a marketing mix for targeted segments. McDonald and Dunbar (McDonald and Dunbar 2004) identified five essential steps from a high strategic segmentation level. Because this dissertation focuses on a new conceptual framework and the unified market segmentation model, the proposed segmentation methodology is more about the new ways of problem definition and how the new solution techniques enable an iterative, interactive and continuing segmentation.

Based on the multiobjective market segmentation definition, the unified market segmentation methodology is depicted in Figure 3.2. There are three loosely coupled components in the methodology: problem definition, multiobjective optimization and solution set analysis.

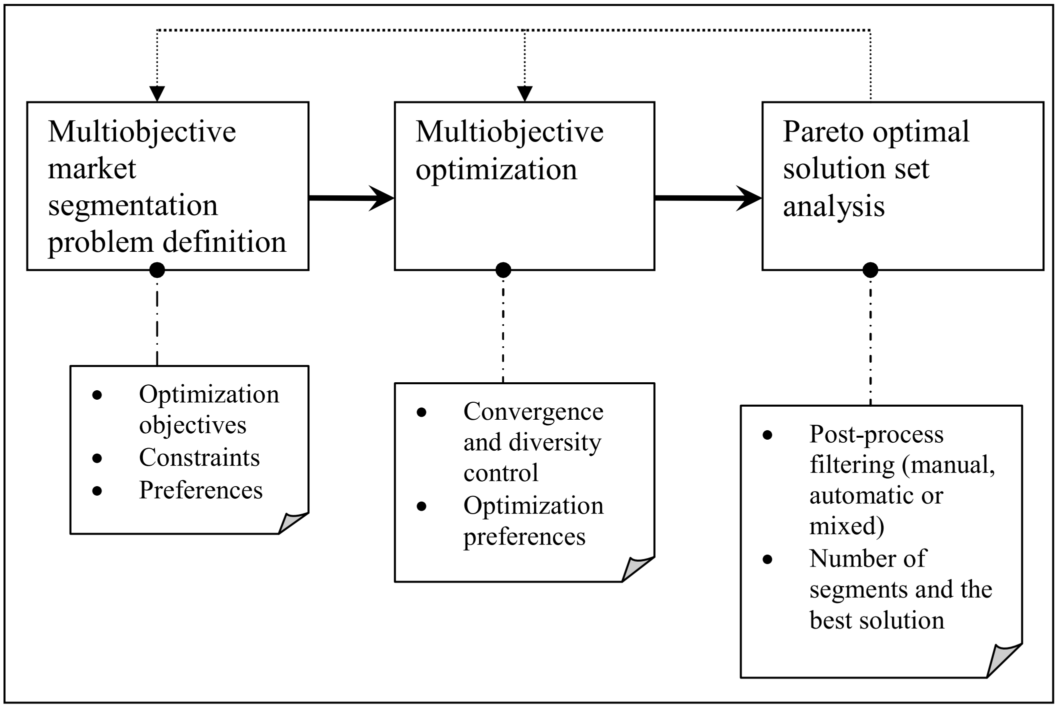


Figure 3.2 The Unified Market Segmentation Methodology

The multiobjective market segmentation problem definition is the action that occurs when one instantiates the general abstract multiobjective optimization problem defined in Section 3.2 with the business-specific optimization objectives. The segmentation objectives are problem specific. A common case is that one objective is segment homogeneity while the other one is a measure of predictive performance for a regression model. In a joint segmentation, they can be within segment variance for each segmentation basis. A single objective segmentation model is a special case here. The optimization objectives often come with managerial, institutional and resource constraints and optimization preferences. The difference between constraint and preference is that constraint is a hard condition put on the possible solutions. For example, one may set the minimum segment size to 30 and the generated solutions will have a size bigger than or equal to 30. Another common constraint will be the number of segments to be generated. This could be a range, such as from 4 to 12, or a single number, like 8. On the other side, the optimization preference is something that is nice to have. An example of preference could be that all segments are of equal size. For a 5 segment solution of 5000 customers, 1000 customers in each segment is ideal but 900 or 1200 customers in one segment is acceptable. The optimization preferences are any heuristic rules/directions that will be incorporated in multiobjective optimization algorithms to influence the generation of a Pareto optimal solution set. Once specified, the optimization algorithm will award solutions with preferred properties and the generated solution set will have more solutions with those properties than solutions without those properties.

The multiobjective optimization component refers to any algorithm that could take the multiobjective problem definition and generate a set of Pareto optimal solutions. The multiobjective methods could be statistical, numerical or meta-heuristic and could be independent of the problem definition and solution set analysis. In multiobjective optimization algorithms, the convergence speed and solution set diversity are the two most important measures. They are antagonistic objectives for the optimization algorithm itself (Ishibuchi and Shibata 2004). Figure 3.3 depicts the conflicts between these two desired properties in a two objective maximization problem. Each star is a Pareto optimal solution found so far by a searching process. From the current solution set, the algorithm tries to move on the A direction because it leads to the Pareto front. On the other side, directions B and C represent the orthogonal diversity direction that is desired by decision makers to have the holistic view of the possible solution space. The diversity measures the representativeness of the generated solution set. The optimization algorithm should be able to incorporate user preference and generate the solution set with the desired properties. As a general case, user preference could be specified up-front as optimization input parameters or during the optimization process interactively. A common case is that users may customize the searching parameters to fine tune the speed of convergence and quality of diversity.

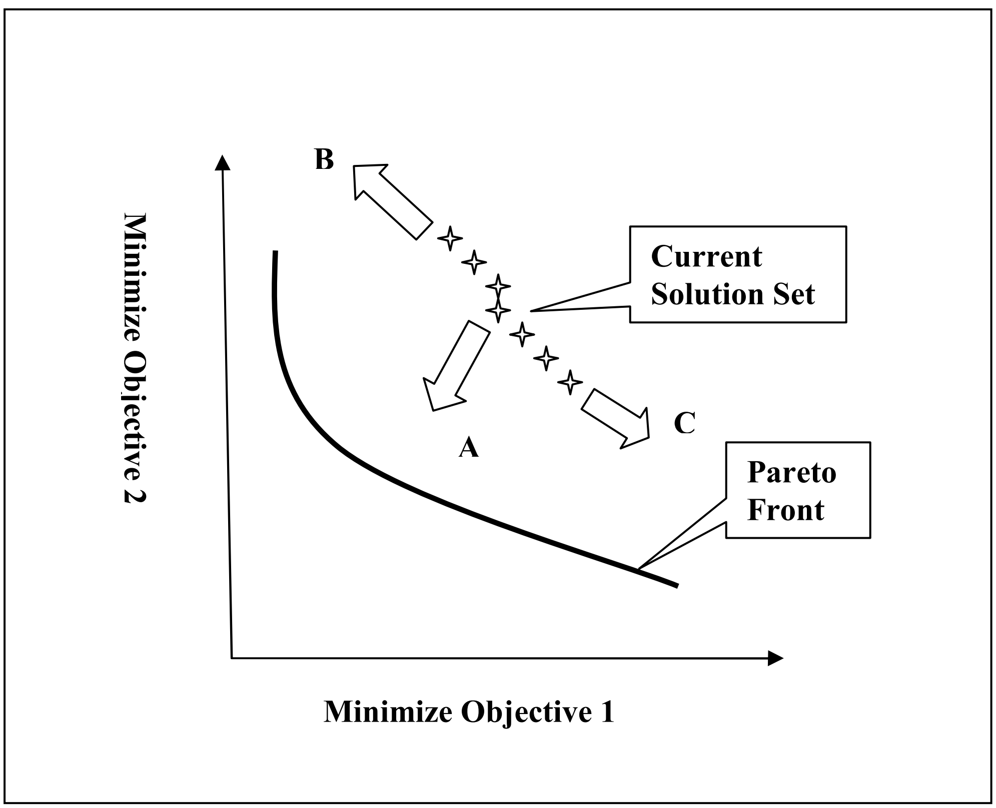


Figure 3.3 Diversity vs. Convergence in Multiobjective Optimization

The third component is the Pareto optimal solution set analysis. The multiobjective optimization problem puzzles decision makers because there are many conflicting objectives and a large number of acceptable solutions. The analysis component is independent of the other two components conceptually though actual implementation could be tightly coupled (Phelps and Koksalan 2003). The analysis techniques are beyond the scope of this dissertation. Some automatic tools can be found in (Mattson, Mullur et al. 2004). As a general rule of thumb, the methodology encourages post hoc constraints and preferences, at least at the early iterations of segmentation, because 1) the up-front constraints and preferences are usually subjective and difficult to define without seeing the possible solutions first; 2) too many subjective constraints and preferences may put unnecessary limitations on search processing and affect the optimization process in unpredictable ways. However, up-front constraints and preferences are preferred for algorithm efficiency and effectiveness because constraints reduce the search space and preferences help to generate biased solution set with desired properties (Padmanabhan and Tuzhilin 2003).

Because the Pareto optimal solution set gives a holistic view of the possible options for the segmentation problem, with the help of visualization tools and data warehouse reporting and analysis tools, decision makers can interact with the solution, look at the profile of each solution, and set filter criteria to focus on interesting solutions. Additionally, the characteristics of Pareto fronts (there is one Pareto front for each number-of-segment value) could be used to automatically suggest the candidate solution.

Objective values of different numbers of segments should be considered separately because they depend on the number-of-segments. The within segment heterogeneity decreases systematically when the number of segments increases. The whole set of customers has the biggest within segment variance that is equal to the total sum of square of the whole set whereas one customer a segment has a within segment variance of zero. Residual variance of linear regression or logistic regression has the similar property. A difficult problem in any market segmentation is determining the number of segments. Unless the problem environment puts a very clear constraint on the number of segments, this is still an unsolved problem in theory and practically many heuristic and statistical methods are used in clustering algorithms (Fraley and Raftery 1998; Tibshirani, Walther et al. 2000; Kleinberg 2002). As is shown by (Handl and Knowles 2004), the Pareto fronts of a range of number-of-cluster values could be used to suggest the number of clusters of the data set for a clustering problem. Though market segmentation is a more complex problem than clustering, the characteristics of Pareto front can be used to suggest candidate values for number of segments.

3.4 An Iterative Segmentation Process

As is shown in Figure 3.2, the methodology suggests an iterative process to define the problem, search the solution space and analyze the results. The findings in each iteration could be used to change constraints, preferences or even redefine the segmentation problem to better serve the segmentation purpose under consideration. The methodology only loosely connects the three components in a segmentation process and each component is independent of the others. Nonetheless, the three components are derived from the formal definition of multiobjective market segmentation described in Section 3.2.

Hoch (Hoch 1994) suggested that combining the adaptability and intuition of domain experts with the comprehensiveness and consistency of a database model gives the best results. The multiobjective optimization approach and Pareto optimal solution set facilitate a new type segmentation methodology that promotes interaction and iteration in market segmentation. It can be described from two perspectives: the segmentation model and business strategy.

The unified market segmentation model could be drawn as shown in Figure 3.4. Marketers interact with all three components of the unified market segmentation model in an iterative process. Both the human expert knowledge and market information (usually consists of CRM, ERP and SCM) play important roles in the entire market segmentation process. Market segmentation usually starts with the problem definition in which marketers translate business goals and needs into multiple optimization objectives and constraints. Soft constraints are defined as preferences to give hints of desired properties of solutions. The problem definition is the input of a specific solution technique that searches and generates a Pareto solution set for the problem. Marketers could control the searching process by setting the runtime parameters of convergence and diversity. These parameters and optimization preferences are usually technique-specific and need insight knowledge of the algorithm and the business domain. The solution technique performance is measured by convergence speed and representativeness of the solution set generation. These performance measures are valuable feedback for problem definition. For example, if the searching is too slow in a large solution space, marketers may want to define more constraints or refine existing constraints to improve the performance. When a set of Pareto optimal solutions are generated, marketers can visualize and/or analyze the solutions from multiple dimensions. Because of the intrinsic complexity of the segmentation problem, the first run may generate less-than-satisfactory results. However, the Pareto optimal set consists of solutions that are distributed along the objective space and are with different number-of-segments. The shape of the Pareto front and the explicit tradeoffs among solutions give useful hints for the good solution. Based on the information of the initial solution set, marketers may refine the constraints, preferences, and/or the search parameters to generate solutions with desired properties. The point here is that each solution set gives more and more insights to marketers regarding the possible segment solutions and their properties. Human expert decisions could be made from the objective results of the data model. This has a big advantage over the up-front weighting of different objectives. This iterative process can occur over and over until a good segmentation is reached.

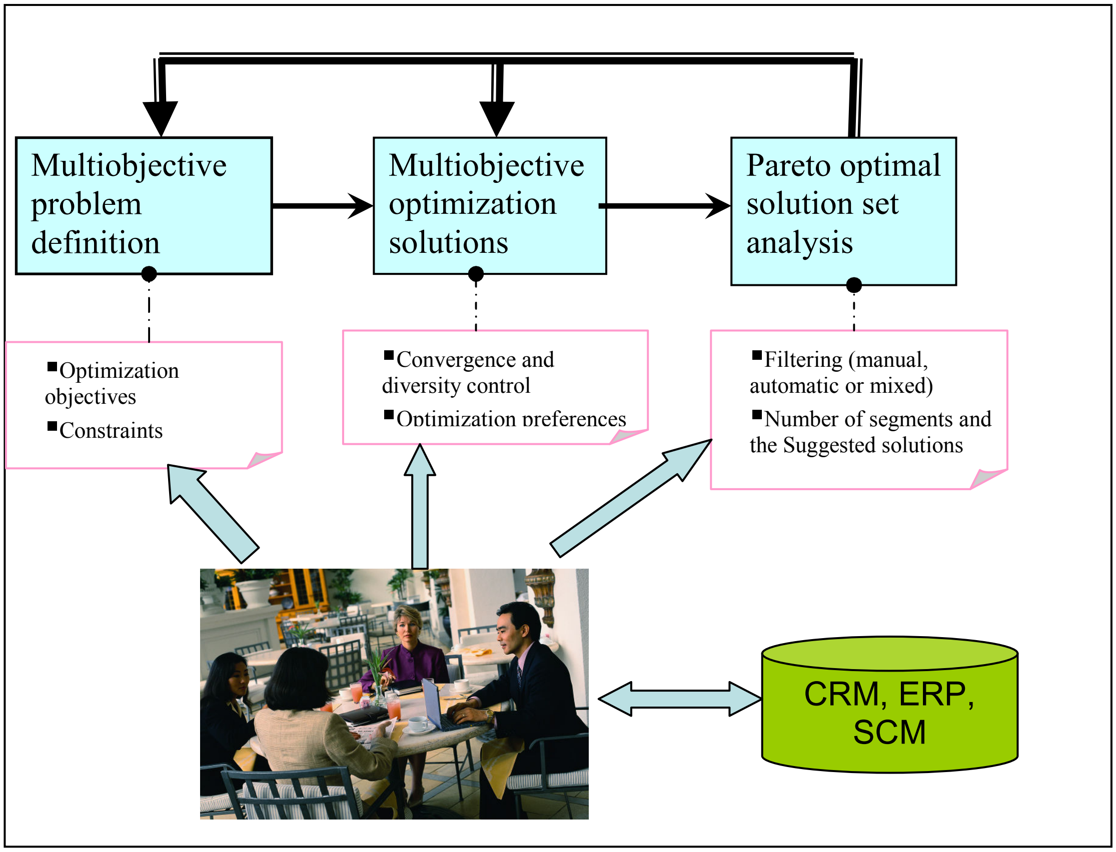


Figure 3.4 The Iterative Interactive Market Segmentation

At the business strategy level, most existing market segmentation methodologies are sequential trial and error approaches because the single objective optimization requires many upfront decisions on parameters and only one solution with an unknown tradeoff is generated. For example, Myers (Myers 1996), Wind (Wind 1978) and McDonald and Dunbar (McDonald and Dunbar 2004) described such methodologies. Though not impossible, it is difficult to work in an iterative way using the single objective optimization method and single optimal solution with unknown tradeoff. The reason is that without a holistic view of solutions in a multiobjective space, a fine tuning of a parameter or constraint may change the segmentation problem definition and its solution significantly. The unified market segmentation model promotes and facilitates multiple segmentation bases and/or multiple criteria. The segmentation process is a continued one in the sense that new customer and marketing data can be incorporated into the problem definition easily in a new iteration. The unified market segmentation allows each objective to be defined independently. The optimization component will simultaneously search solutions with regard to all objectives. The generated Pareto optimal solutions will show the tradeoffs in each objective/criterion explicitly in the solution space.

4 MULTIOBJECTIVE MARKET SEGMENTATION USING EVOLUTIONARY ALGORITHMS (MMSEA)

Market segmentation is NP-hard in computational complexity (Krieger and Green 1996). Usually efficient heuristic methods are required to address this intrinsic complexity. As many problems are multiobjective in nature, considerable effort has been devoted to solving them. Before the 1950s, limited by the mathematical techniques and computation capabilities, multiobjective problems were usually converted into a single criterion problem and then addressed by global optimal methods such as greedy algorithm and Hill-Climbing. Kuhn and Tucker (Kuhn and Tucker 1951) first introduced the "vector maximum problem" that directly targets multiobjective problems. Goal programming (Charnes and Cooper 1961) was a popular method for many years. However, the complexity of multiobjective problems requires more efficient techniques. A variety of stochastic search strategies such as evolutionary algorithms (Goldberg 1989; Back 1996), tabu search (Glover and Laguna 2004), simulated annealing (Metropolis, Rosenbluth et al. 1953; Kirkpatrick, Gelatt Jr. et al. 1983), ant colony optimization (Dorigo, Maniezzo et al. 1996) and memetic (Moscato 1999) algorithm have been developed. Among these approaches, evolutionary algorithm has a long proven history. The survey by Coello et al (Coello, Veldhuizen et al. 2002) showed that the evolutionary algorithm is the most widely used meta-heuristic approach to solve multiobjective optimization problems.

4.1 Multiobjective Evolutionary Algorithms (MOEA)

The multiobjective evolutionary algorithm (MOEA) was initially proposed by Rosenberg (Rosenberg 1967). The first implementation was suggested by Schaffer (Schaffer 1984). Since then, evolutionary multiobjective optimization algorithms (Coello, Veldhuizen et al. 2002) have developed rapidly for multiobjective problems. The characteristics of MOEA make it particularly suitable for market segmentation problems.

- It searches for optimal solution(s) using a set of objectives simultaneously. This property makes the algorithm efficient and globally optimal.
- The evolutionary algorithm can be used to find a set of solutions that has the desired diversity. The diversity requirement could avoid the local optimum problem that often is common in other approaches. The final results are usually representative of the possible tradeoffs that are important for decision making.
- It is less susceptible to the shape of continuity of the Pareto solution set because it does not make any assumptions about data and problem properties.
- It addresses both searching and multiobjective decision making at the same time.
- Usually the search algorithm is independent of objective functions and decision variables. This is a very attractive feature because the algorithm could be used in a broad range of problems.

4.2 MMSEA

Based on the proposed unified segmentation model, we developed the program of multiobjective market segmentation using evolutionary algorithms (MMSEA) as the multiobjective optimization component. The high level components of the algorithm are shown in Figure 4.1. As will be demonstrated by our evaluation in the following chapters, the quality of the initial solution set is very important. The initialization component is developed for generating high quality initial solution sets for different optimization objectives. Parent selection, crossover, mutation and archive are typical components of evolutionary algorithms. In evolutionary algorithms, each solution candidate is called an individual. The set of solution candidates is called the population. The archive component and memory components are used to select individuals for crossover (recombination) and store promising individuals. The stop criterion could be that either a number of generations have been run or the Pareto front progress is under a threshold. Because clustering algorithms are often used in market segmentation, a segment is often called a cluster in the following discussion.

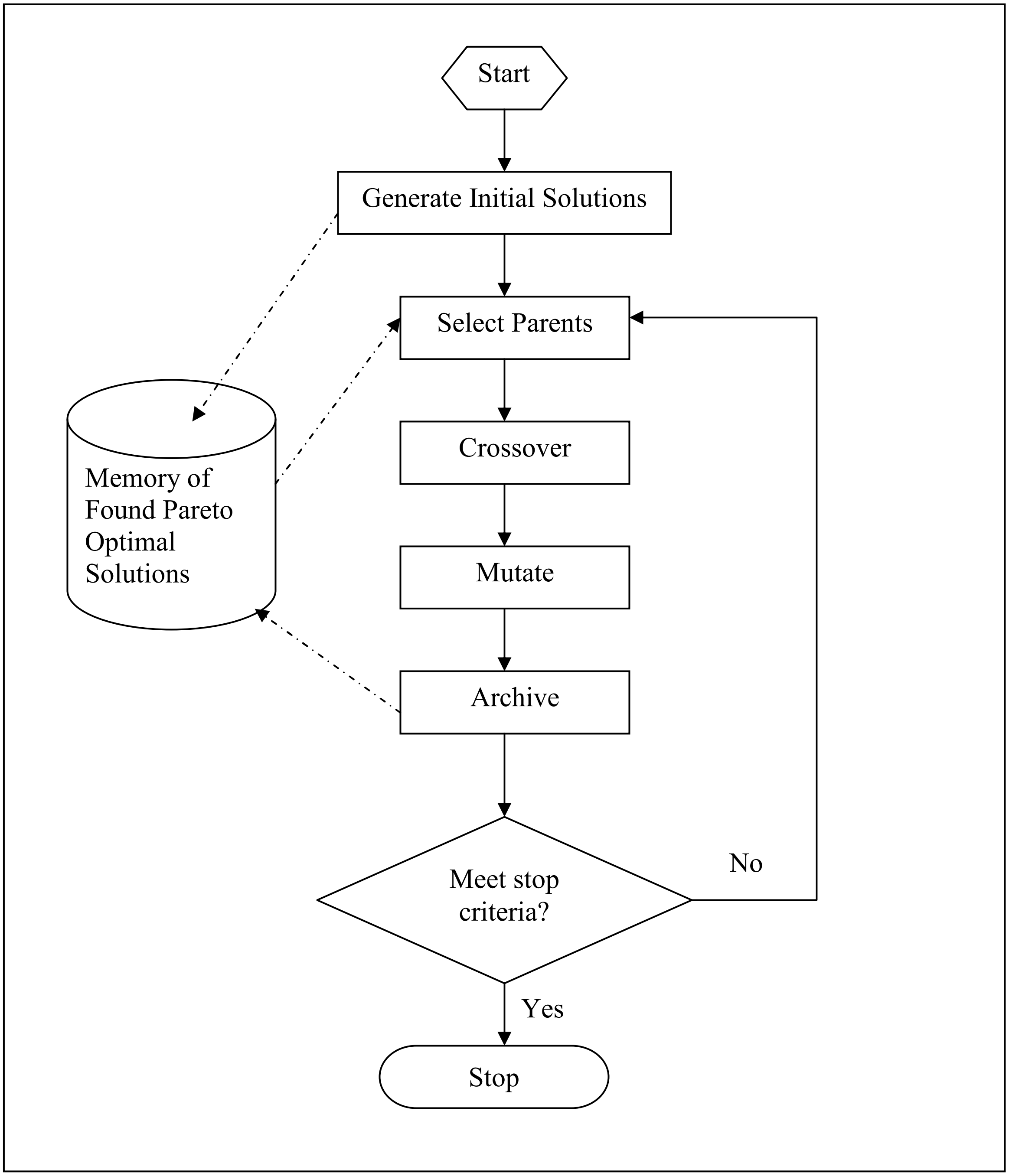


Figure 4.1 Components of MMSEA algorithm

Unlike a single objective optimization algorithm, a multiobjective optimization algorithm has to remember a set of solution candidates. The search process itself is a multicriterion problem because it has two basic tasks. One task is to find the best Pareto optimal solution set. The second task is to find a set of solutions that are representative of possible solutions, thus allowing decision makers to analyze a set of diverse solutions. We will discuss important algorithm design decisions in the following sections.

4.2.1 Gene Representation

Maybe one of the most important decisions is the representation of genes. Intuitively one can assign a segment number to each customer. However, a big disadvantage of such representation is that both the crossover and mutation operations can only be applied on the individuals with the same number of segments. MMSEA is designed to generate a number of solutions of different number of segments. Generally, a good 3-segment solution could be a parent for a good 4-segment solution and the reverse is often true. Park and Song (Park and Song 1998) proposed the graph-based adjacency representation that has the advantage of allowing solutions with different number of segments to mate and mutate. MMSEA uses this approach to encode the cluster membership. Each customer has a gene position in the chromosome. The allele value is the position of a gene that is in the same segment. For a single element segment, the allele value is itself.

4.2.2 Initial Solution Set Generation

The quality of the initial solution set has a big impact in the performance of evolutionary algorithms. A simple naïve initialization approach is random assignment of customers, but the simple single objective optimization methods often fit the initialization task much better. The different optimization objectives require different initialization algorithms to generate the corresponding initial solution set.

For the basic task of clustering, the optimization objective is to minimize the within segment heterogeneity. Handl and Knowles (Handl and Knowles 2004) showed that it is good to generate the initial solution set from two complementary methods: minimum spanning tree (MST) and K-means. Each method generates 50% of the initial solution set for the clustering objective. To improve the speed of convergence of the algorithm, the initial MST encoding is also used to encode the K-means output, i.e., for any two genes in the same cluster in K-means output, if a link is found in the initial MST, the allele value of the MST gene is copied as the value of the corresponding K-means gene.

The responsiveness criterion is often met by using a regression model. Clusterwise regression (Spath 1979; Spath 1982) is used to initialize solutions for regression objectives. Java doesn't have good support for statistical functions. We chose to use R statistical environment (R Development Core Team 2007) to perform statistical functions. There are two approaches for an application to call R functions. One is to use

Rserve (Urbanek 2003). Rserve allows applications to call R functions through TCP/IP. The advantage is that all calls are made in remote mode without the need to initialize R and link to the R library. However, the performance is not good if the data size is large. Because MMSEA calls R many times and a large amount of data needs to be shared between R and Java, a Java/R interface (JRI) (Urbanek 2006) is used. JRI allows Java applications to call R methods in a different thread. Because R and Java share a single process space, the performance is much better compared to the Rserve approach.

Additionally, both K-means and clusterwise regression only optimized for one objective. To fill the gap between the two or more objectives, we developed a multistage algorithm to generate initial solutions. Clusterwise regression followed by K-means and K-means followed by clusterwise regression methods are used to generate solutions that are optimized for different objectives in different stages.

Figure 4.2 shows the evolution of solutions with random initialization. Figure 4.3 shows the evolution of solutions with clusterwise and K-means initialization. At 1,000 generations and 20,000 generations, the clusterwise and K-means initialization generates a much better Pareto front in terms of optimization objective values and solution diversity.

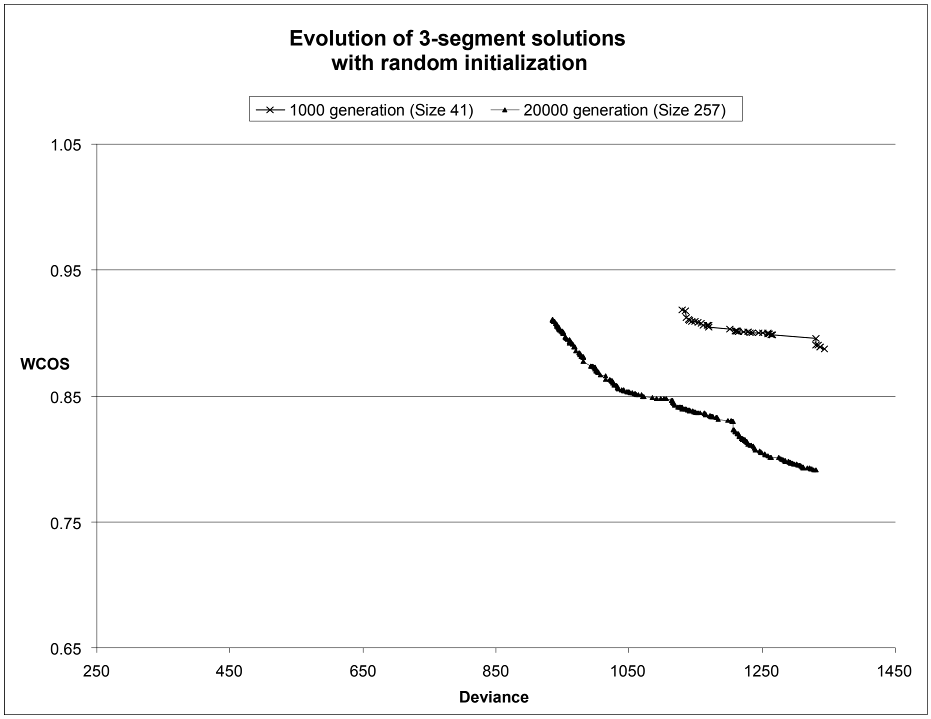


Figure 4.2 Solution Evolution with Random Initialization

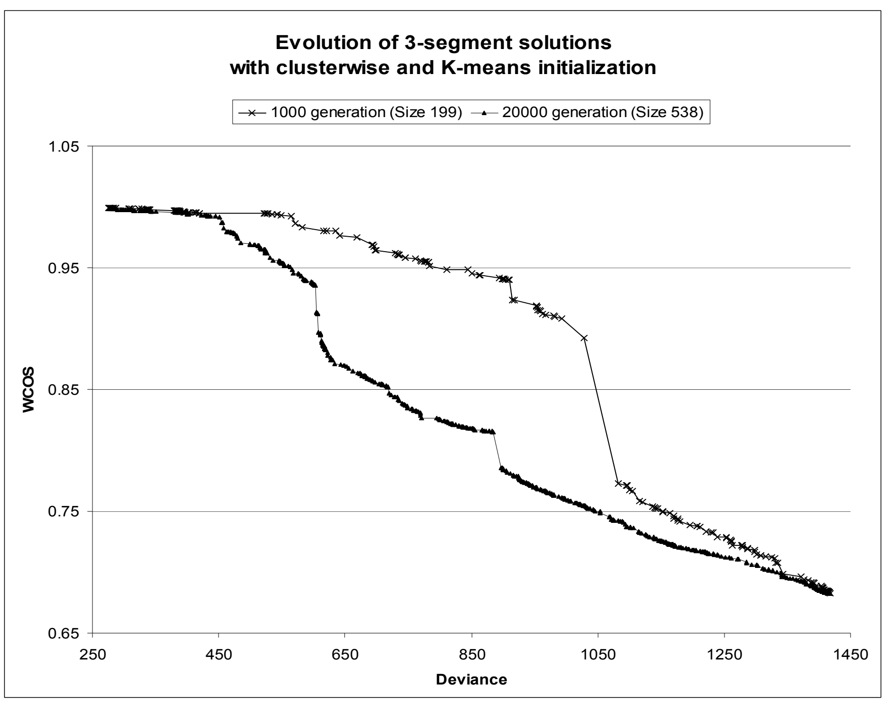


Figure 4.3 Solution Evolution with clusterwise and K-means Initialization

4.2.3 Parent Selection

A three-tier selection algorithm is developed to select the parents for reproduction. The first tier is the number-of-clusters. For each number-of-clusters value, there exists a solution space whose dimensions are the segmentation objectives. The solution space forms the second tier. The third tier is the hyperbox in the solution space. The hyperbox is occupied by zero, one, or more solutions whose values fall within the boundary of the hyperbox. MMSEA first selects a number-of-clusters value from the range of number of clusters randomly. Then, in the corresponding solution space, MMSEA selects two occupied (with at least one solution) hyperboxes randomly. In the third step, it selects one solution randomly from each hyperbox. Binary tournament is used to select the final winner whose hyperbox is least crowded. This method helps select diverse solutions for each value of number-of-clusters. It allows crossover between two solutions with different values of number-of-clusters.

The three-tier algorithm helps to find parents for the diversity purpose. However, the MMSEA needs to move to the Pareto front as fast as possible. So in the binary tournament selection, an ideal-based strategy is developed to select solutions that near the ideal solution. An ideal solution is a solution that has the minimum values with regard to all objectives. It's ideal in the sense that practically it is impossible to reach this solution.

4.2.4 Crossover and Mutation

Two parents are selected for uniform crossover. During crossover, each gene has a chance of 0.5 (the default configuration) to swap with a corresponding gene in another parent.

The valid allele value of each gene is the position of any gene, even itself. A random mutation with this large number of choices could be very slow. We used the concept of neighbor scope introduced by MOCK (Handl and Knowles, 2004): the possible mutation value is within the L nearest neighbors. Unlike MOCK that uses a fixed value of L, MMSEA uses a dynamic value that is calculated by dividing the length of the chromosome by the number-of-clusters of the mutating chromosome.

4.2.5 Solution Archive

An elitist mechanism similar to PESA-II (Corne, Jerram et al. 2001) is used as the fitness sharing control. Selected parent solutions are called internal populations in PESA-II. Internal populations, after crossover and mutation, are to be placed in an external population that stores all the non-dominated solutions found so far. In addition to the standard dominating comparison, MMSEA requires that each new solution is a minimum distance away from any existing solution. This modification is important because it deals with precision loss in calculation and improves diversity of the final solution set. Otherwise, two solutions with the same value or of little difference will be archived.

4.3 Solution Suggestion and the Number-of-cluster Detection

MMSEA generates a set of Pareto optimal solutions. The Pareto front characteristics make some solutions more interesting than the others. We have developed a maximum distance algorithm for the unified market segmentation model to suggest the "best" solution for each number-of-segments value. If the solutions are evenly distributed in the solution space, an angle-based algorithm (Branke, Deb et al. 2004; Mattson, Mullur et al. 2004) could be used to find the knee of the Pareto front. Figure 4.4 shows the possible shape for a two-objective Pareto front with 9 solutions that are not distributed evenly in the solution space. The angle-based approach (Branke, Deb et al. 2004; Mattson, Mullur et al. 2004) uses the angle formed by one point and its (two or more) neighbors to find the knee of the Pareto front. The knee is the point that has the biggest angle. As illustrated by the upper-left part of Figure 4.5, the algorithm fails to find the global optimal knee because of the local maximum formed by the three points clustered together.

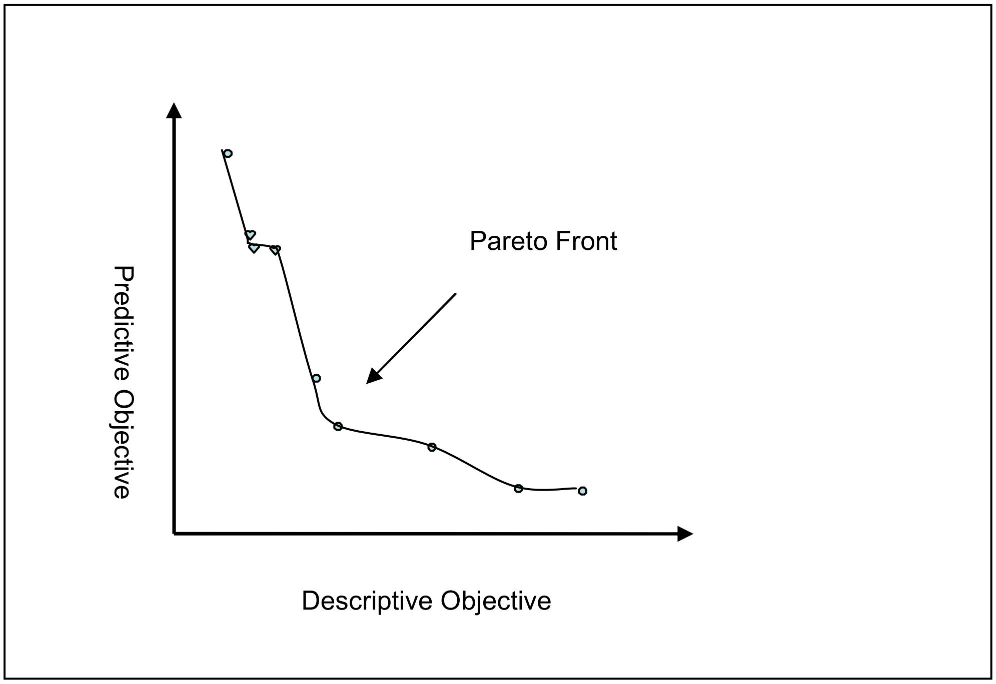


Figure 4.4 The Pareto Front of a Solution Set

The maximum distance algorithm is proposed to find the knee of the Pareto front whose solutions are distributed with different densities in the solution space. If we assume that the two extreme solutions form an indifference line, i.e., solutions in this line are equally good, solutions sitting below this indifference line are better solutions because they represent a set of solutions that have better gain/loss rate. Then the distance from each intermediate solution to the line is calculated and the solution that (i) has the maximum distance and (ii) is below the reference line is the suggested solution for this Pareto front. Figure 4.5 depicts the process of finding maximum distance solution.

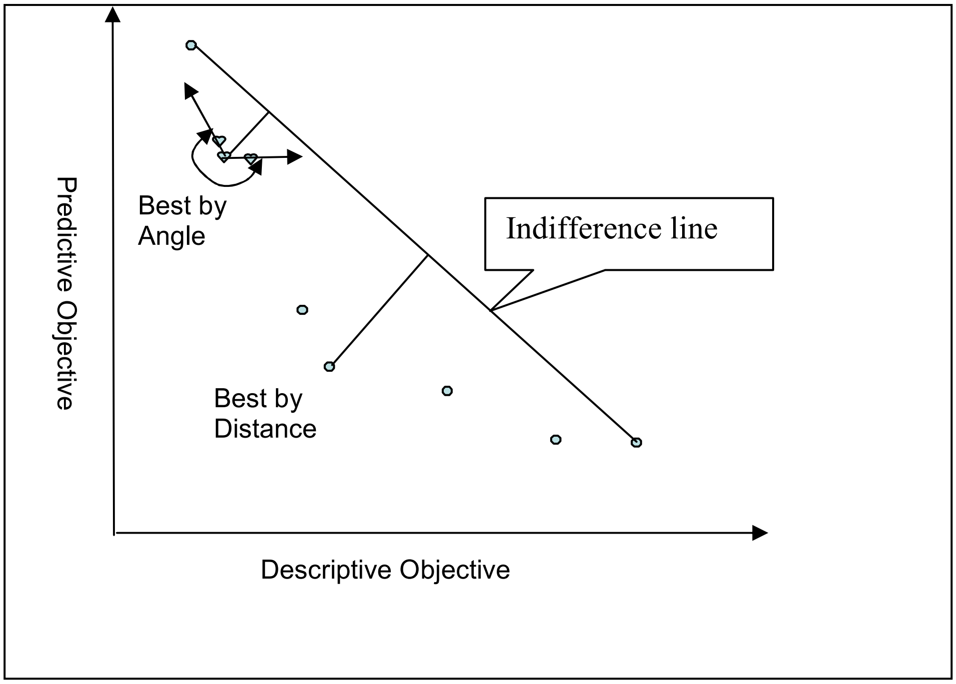


Figure 4.5 Finding Knee of Pareto Front

The angle-based algorithm can be relied on to suggest the value of number-of-segments, as is shown in Figure 4.6, because of equal distances between two consecutive number-of-segments values. Consequently there are two steps in finding the "best" solution. In the first step, MMSEA finds one solution for each number-of-segments value using the maximum distance algorithm. In the second step, MMSEA calculates the angle (α value in Figure 4.6) for each objective of the solution and adds them up for each solution. The solution that has the biggest total angle value is the suggested "best" solution based on the characteristics of the solution set.

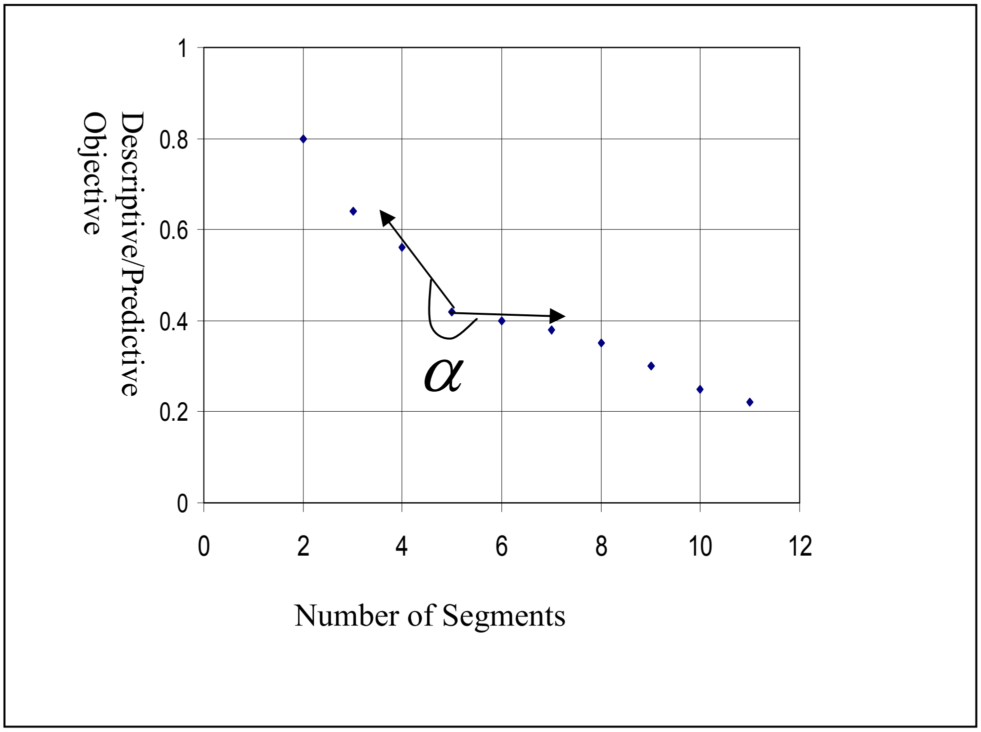


Figure 4.6 Suggest a Good Value of Number of Segments

5 EVALUATION ONE: JOINT DESCRIPTIVE MARKET SEGMENTATION

To justify the benefits of the unified segmentation model and its implementation discussed in the previous chapters, we need to apply it in different business settings. Ideally, the empirical evaluation should cover both descriptive and predictive market segmentation. This chapter describes the empirical evaluation of joint market segmentation. The following two chapters discuss the empirical evaluation of linear and logistic predictive market segmentation. All is done using real business data with a multicriterion segmentation model.

The joint descriptive market segmentation is used as the first evaluation case for its conceptual simplicity. The same optimization function is applied in two segmentation bases. The comparison with the well-known K-means algorithm shows the difference between the proposed multiobjective solution technique and the existing single objective optimization methods. Following chapters will compare the new unified market segmentation model with the state of the art segmentation methods that are based on more complex statistical models.

5.1 Segmentation Model

Vargo and Lusch (Vargo and Lusch 2004) as well as Boulding et al. (Boulding, Staelin et al. 2005) viewed marketing and CRM as management of the dual creation of firm (shareholder) and customer values. Customers choose between competing products and services based on their assessment of superior value: in other words, they choose the proposition that consists of the benefits they are looking for at a price they perceive as providing superior value for the money (McDonald and Dunbar 2004). The challenge for every company is to be able to understand and differentiate heterogeneous customers by their needs and their responses to marketing mix variables to deliver the winning propositions profitably. Accurate market segmentation can significantly improve targeting of marketing campaigns and directly contribute to a business's bottom line. The segmentation goal of this research is to understand the value proposition of the cell phone service market. The wireless service industry has a high churn rate ranging from 23.4 to 46% annually. The bi-directional value proposition-based joint market segmentation is useful for many marketing activities. The goals of those activities could be improved customer loyalty, win-win product and service positioning, proactive customer retention and targeted marketing promotion. From the view of service providers, the customer value is measured by customer revenue and service cost (phone usage and number of customer service calls). From the customer's perspective, the surrogates of customer benefits are minutes of phone call, number of phone calls, and quality of services measures such as number of blocked calls and number of dropped calls.

In a simplified view, the customer value and customer benefit could be depicted by the value proposition matrix that is depicted in Figure 5.1. Based on the user usage pattern and profit pattern, a segment-based marketing campaign could be launched to retain high profit users and/or to convert low profit users to high profit segments. Elliott and Glynn (Elliott and Glynn 2000) proposed the loyalty and value segmentation model for industrial buyers. Compared with their model, the proposed value proposition model has a big advantage when used in the consumer service market: the customer benefit is easier to measure than customer loyalty because the benefit could be directly measured by the service usage. Often high customer loyalty is a result of a long time high customer benefit. Customer benefit is more dynamic than customer loyalty because it changes in every interaction between the customer and the firm. Bock and Uncles (Bock and Uncles 2002) found that profit is often more cost-effective than other segmentation bases because of its dynamic nature. Nonetheless, customer benefit can be a good proxy of customer loyalty.

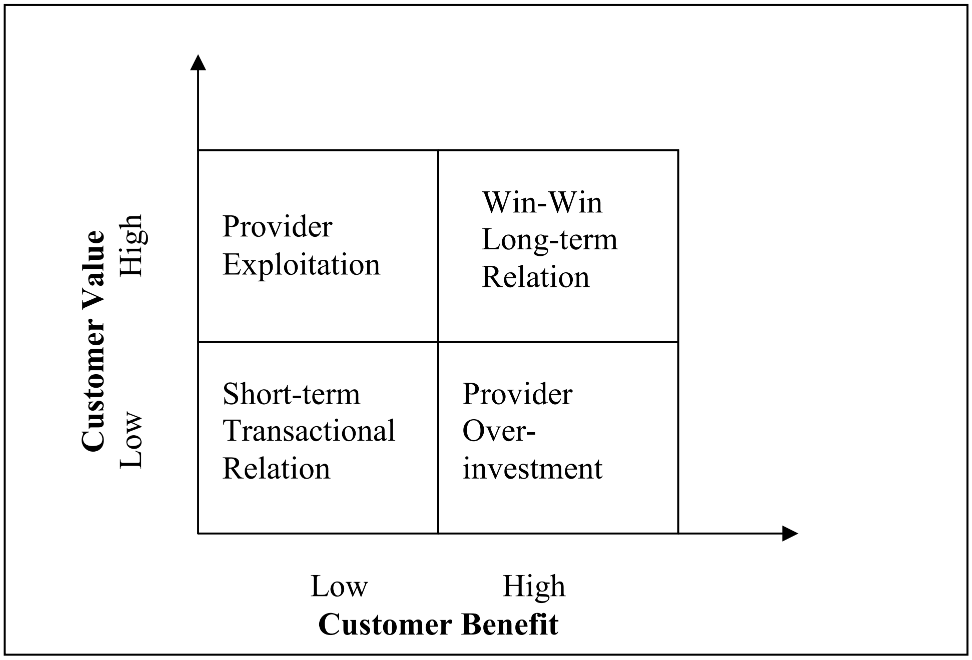


Figure 5.1 Customer Value Proposition Model

In this segmentation matrix it should be noted that the only path to a long-term customer relationship is by the firm receiving high value from the customer and vice versa i.e., the customer receiving high benefit from the firm (i.e., the upper right cell of the matrix). In the case of low customer value to the firm and low customer benefit, both parties will discontinue the relationship unless value for both can be increased (i.e., lower left cell of the matrix). On the other hand if the benefits are asymmetric (see the off-diagonal) then either the customer will be motivated to switch, for example, when the value to the firm is high and the customer benefit is low (i.e., the upper left cell of the matrix); or the firm will be motivated to increase fees or lower service when the customer receives high benefit but the firm is not benefiting in terms of value received (i.e., lower right cell of the matrix). Clearly this type of segmentation strategy not only provides homogeneous groups of customers for targeting marketing efforts but also provides strategic guidance for each of the four segmentation groups as outlined in the preceding matrix.

5.1.1 What's Wrong with Current Segmentation Methods

Traditional segmentation methods optimize only one objective, for example, the minimization of within group heterogeneity for customer benefit. Those methods may generate a solution as shown in Figure 5.2.

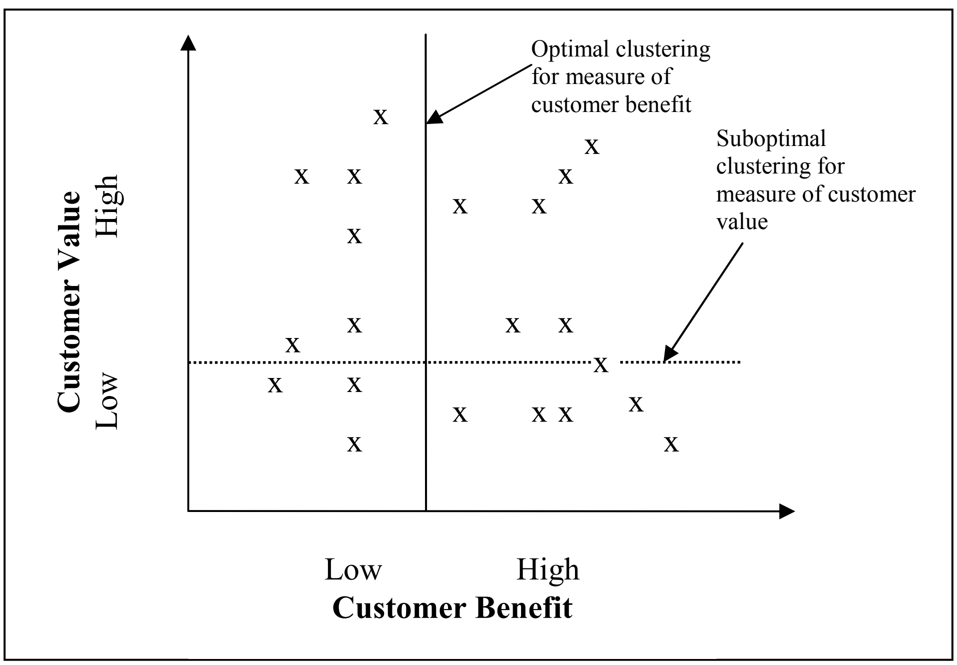


Figure 5.2 Segmentation Optimized for Customer Benefit Objective

In this case, marketing campaigns that are based on the suboptimal segmentation may incorrectly target some low value customers. Consequently, the marketing campaign is not as effective as it could be.

The fundamental reason behind the suboptimal solution is that the measure of customer value and the measure of customer benefit do not agree with each other. Technically, the optimized within group heterogeneity measures for customer benefit and customer value are located in separated positions as shown in Figure 5.3.

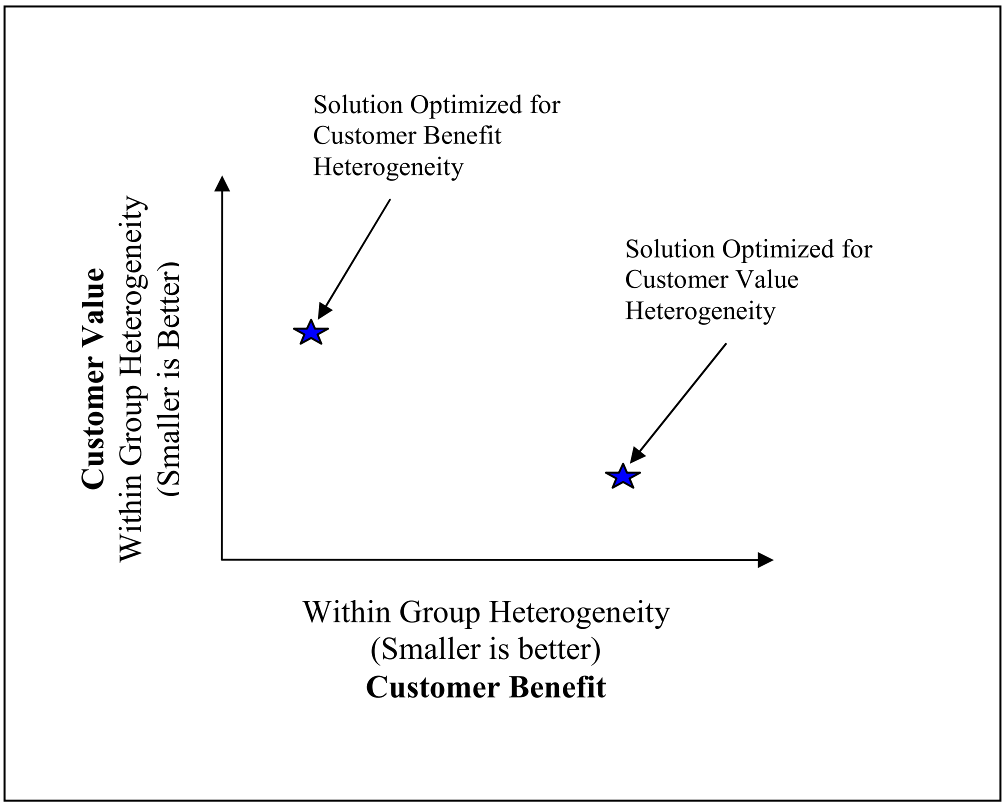


Figure 5.3 Antagonistic Objectives

5.1.2 Multiobjective Method

The proposed multiobjective method optimize within group homogeneity for both customer value and customer benefit. It may generate a global optimal solution as depicted by Figure 5.4.

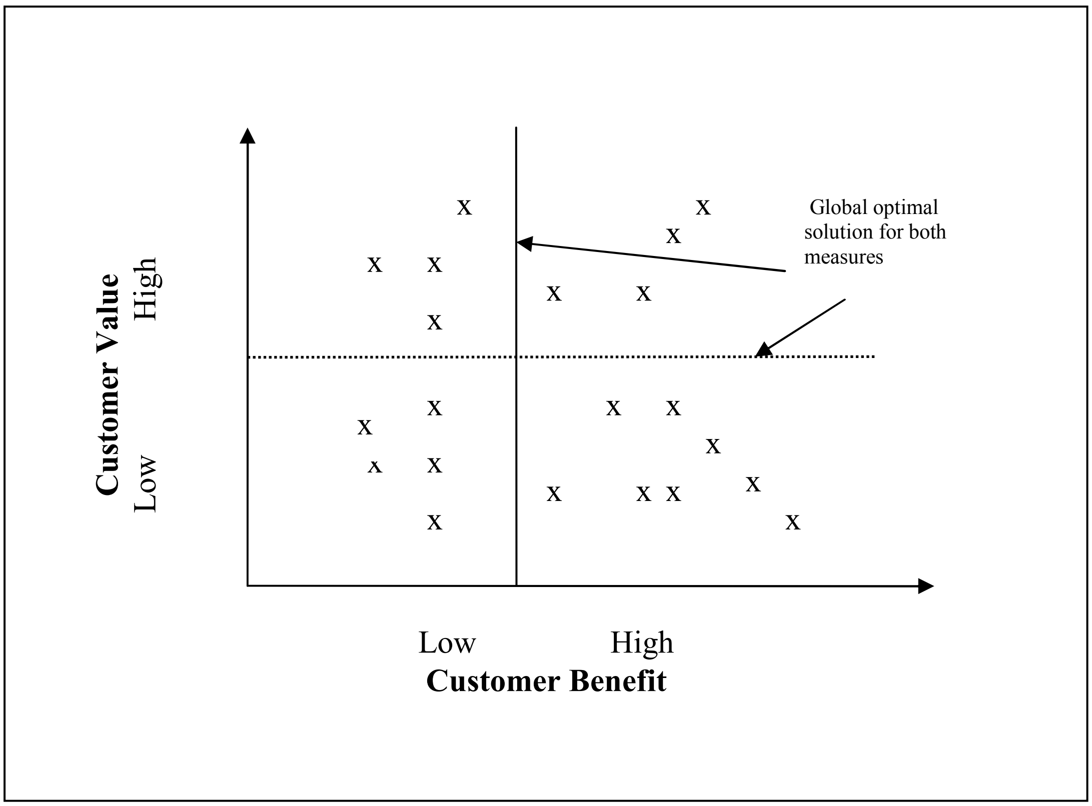


Figure 5.4 Multiobjective Segmentation

As shown in Figure 5.5, MMSEA helps to find the balanced segmentation solutions that are Pareto optimized for all objectives. Another distinct advantage of our approach is that the algorithm generates multiple Pareto optimal segmentation solutions in a single run. This gives the decision maker much desired flexibility in interpreting and selecting the "right" segmentation solution for the specific segmentation goal. For big customer bases, as is in the case of cell phone service, a 1% improvement in targetability potentially could bring a great amount of profit for the service provider because of the improved customer lifetime value.

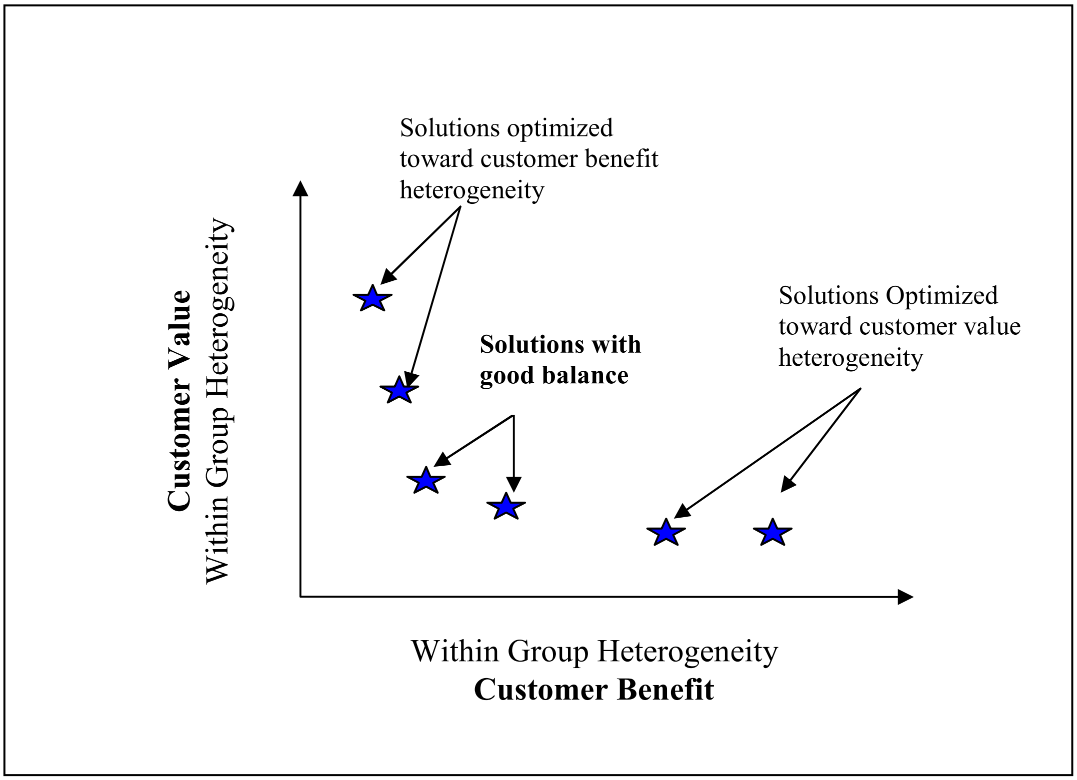


Figure 5.5 Pareto Front of Multiobjective Optimization

5.2 The Data Set and Segmentation Definition

The data set was donated by an anonymous cell phone service provider to the Teradata Center for Customer Relationship Management at Duke University. It consists of a calibration database of 40,000 customers and a validation database of 31,047 customers. The calibration database is over-sampled for churn customers while the validation database is not. In addition to churn flag, each customer has a set of 76 attributes that could be classified into behavioral data and demographic data. The behavioral attributes, as in most CRM database, have a very low missing value rate of 0.24%. After filtering out the missing value cases, we randomly sampled 1,000 cases from 30,972 cases in the validation database. There are total 19 churn customers in our sample. The churn rate is 1.9%. Based on the value proposition segmentation model, 6 attributes are selected as customer benefit basis and 4 attributes are selected as customer value basis. The descriptive statistics are shown in Table 5.1. Consequently, the value proposition segmentation is a joint descriptive segmentation based on two bases.

	Minimum	Maximum	Mean	Std. Deviation
Customer Benefit Attributes				
Mean monthly minutes of use	.00	3830.25	528.66	540.62
Mean overage minutes of use	.00	1317.00	39.19	100.98
Mean number of outbound voice calls	.00	393.33	26.51	36.46
Mean number of inbound voice calls	.00	231.33	8.57	17.15
Mean number of dropped voice calls	.00	98.00	5.92	8.62
Mean number of blocked voice calls	.00	135.00	4.26	10.40
Customer value attributes				
Mean monthly revenue	2.21	526.99	59.03	49.33
Mean total recurring charge	-.08	219.99	47.01	23.30
Months in Service	6	58	18.40	9.97
Mean number of customer care calls	.00	39.33	2.01	4.48

Table 5.1 Statistics of Segmentation Bases

It is worth noting that all values are standardized before analysis because we want to treat all variables equally. The research by Milligan and Cooper (Milligan and Cooper 1988) suggested that standardizing the Euclidean distance similarity measure usually helps.

Though standardization by range is the best method according to their study, we chose the less effective z-score transformation for its good interpretation and good support of statistical software packages.

The within cluster omega squared (WCOS) was used to measure the quality of segmentation for each segmentation basis. It’s defined as follows:

$x_{ij} =$ the value of attribute j for customer i; $i = 1,...,I$, I is the number of customers; $j = 1,...,J$, J is the number of attributes in the segmentation basis;

$I(c) =$ the set of customers in the cluster c;

i. $\bar{x}_{j_c}$ = the mean of attribute j of cluster c: $\bar{x}_{j_c} = \frac{1}{|I(c)|}\sum_{i \in I(c)} x_{ij}$;

ii. $\bar{x}_j$ = the mean of attribute j for all customers: $\bar{x}_j = \frac{1}{I}\sum_{i=1}^{I} x_{ij}$;

iii. Within Cluster Sum of Squares (WCSS) $= \sum_{c=1}^{K}\sum_{j=1}^{J}\sum_{i \in I(c)}\left(x_{ij} - \bar{x}_{j_c}\right)^2$;

iv. Total Sum of Squares (TSS) = $\sum_{j=1}^{J}\sum_{i=1}^{I}\left(x_{ij} - \bar{x}_j\right)^2$;

v. Within Cluster Omega Squared (WCOS) $= \frac{WCSS}{TSS}$;

As an instance of the general multicriterion market segmentation problem defined in Section 3.2, the segmentation problem is defined as following:

z_i = the segment membership of customer i; $i = 1,...,I$, I is the number of customers. There are 1000 sample customers in this study.

$z = [z_1, z_2, ..., z_I]$ is a vector of decision variables that represent a segmentation solution for all I customers;

(1) $F(z) = [f_1(z), f_2(z)]$ is the objective vector to be optimized;

- $f_1(z)$ = minimize the WCOS of customer benefit basis;
- $f_2(z)$ = minimize the WCOS of customer value basis;

(2) $G(Z) = [g_1(z)]$ is the constraint vector to be satisfied;

- $g_1(z)$ = the number of segments. The user msut specify one or a range of number of segments

(3) A specific segmentation solution $z^* = [z_1^*, z_2^*, ..., z_I^*]$ is Pareto optimal if there does not exist another solution z^l such that

- z^l satisfies (2) and
- $f_j(z^l) \leq f_j(z^*)$ for all $j = 1,2,...,M$ and
- $f_k(z^l) < f_k(z^*)$ for at least one k

(4) The goal is to find one (or more) segmentation solution(s) $z^* = [z_1^*, z_2^*, ..., z_I^*]$ that is (are) Pareto optimal with regard to the objective vector (1) and satisfies constraint vector (2).

As defined in Section 3.2, $z = [z_1, z_2, ..., z_I]$ is a segmentation solution for all I customers. There is no up-front constraint in this study for two reasons. First, constraints on the search space may prevent MMSEA from finding a globally optimal solution. Second, as will be shown later, constraints can be applied easily after the Pareto solution set is generated. As the value of WCOS decreases with increasing value of number-of-clusters, the dominating comparison is only conducted for solutions that have the same number-of-clusters.

5.3 K-means Results

As a benchmark for MMSEA, K-means clustering with a maximum of 500 iterations was used to segment the customer data set. Because K-means is a single criterion clustering algorithm and is sensitive to initial cluster assignment, K-means clustering was executed 100 times for each segmentation basis with a number-of-clusters ranging from 3 to 11. The best of 100 runs was used as the final result for the segmentation solution. Additionally, K-means was executed 100 times on all customer benefit and customer value attributes (totally 10) and two results were selected for each number-of-cluster value: one had the best WCOS of customer benefit basis and the other had the best WCOS of customer value basis. The results of all solutions are shown in Table 5.2 and Table 5.3 and in Figure 5.6.

	Optimized for Customer Benefit		Optimized for Customer Value	
Number-of-clusters	Customer Benefit WCOS	Customer Value WCOS	Customer Benefit WCOS	Customer Value WCOS
3	0.55	0.84	0.84	0.60
4	0.47	0.83	0.84	0.49
5	0.43	0.79	0.78	0.41
6	0.39	0.78	0.79	0.36
7	0.35	0.79	0.76	0.33
8	0.32	0.78	0.76	0.30
9	0.30	0.77	0.75	0.27
10	0.30	0.76	0.73	0.25
11	0.26	0.76	0.70	0.24

Table 5.2 K-means Results (1)

Optimized for all attributes				
	Best for Customer Benefit		Best for Customer Value	
Number-of-clusters	Customer Benefit WCOS	Customer Value WCOS	Customer Benefit WCOS	Customer Value WCOS
3	0.59	0.72	0.68	0.67
4	0.52	0.71	0.59	0.59
5	0.48	0.67	0.58	0.50
6	0.46	0.59	0.56	0.47
7	0.46	0.49	0.52	0.45
8	0.43	0.46	0.47	0.44
9	0.41	0.46	0.43	0.41
10	0.40	0.43	0.40	0.40
11	0.38	0.42	0.43	0.38

Table 5.3 K-means Results (2)

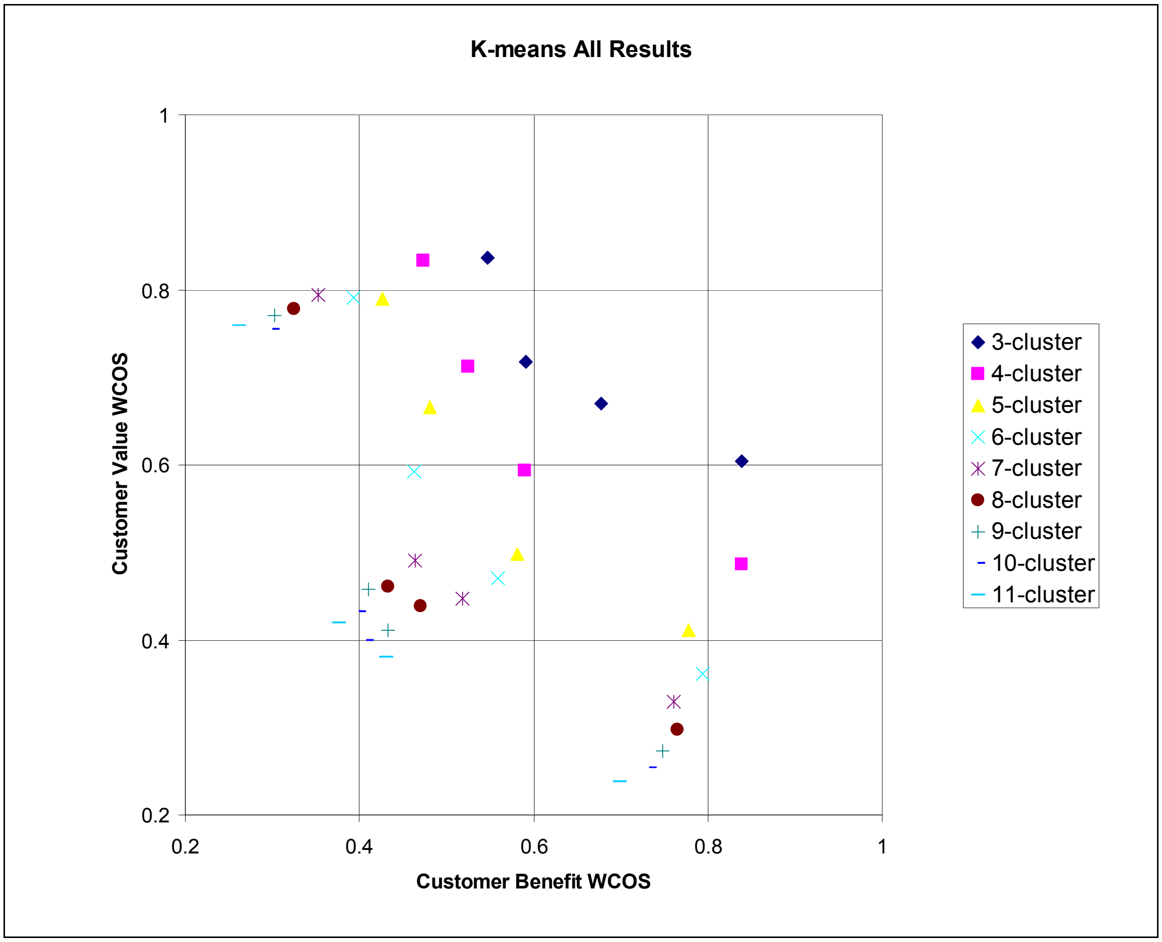


Figure 5.6 Solutions of K-means

Several observations can be made from these results. First, as expected, the customer benefit WCOS and customer value WCOS are antagonistic. The left top and bottom right points are the optimal solutions for each segmentation basis. The intermediate points are results of all-attribute executions that represent "unknown" tradeoffs of multiple criteria. Second, it is hard to find a solution that has adequate tradeoff between the two criteria because the K-means clustering doesn't support optimization on more than one criterion directly. Third, K-means generates one single locally optimal solution though the multicriterion problem has many possible Pareto optimal solutions. These issues are shared by most single objective-based segmentation methods discussed in chapter 2. They lack the flexibility and insights that could be gained from a multiobjective-based method.

5.4 MMSEA Results

MMSEA generates a set of Pareto optimal solutions for a range of number-of-segments values in a single run. Table 5.4 shows the algorithm parameter values of MMSEA.

Parameter Name	Parameter Value
Number of Generation	2500
Internal population Size	100
External population Size	1000
Number-of-segments	Min: 2 Max: 10
Crossover Rate	0.5
Mutation Rate	Dynamic

Table 5.4 MMSEA Parameter Settings

The parameters are derived from user-specified external population size and number-of-segments. Practically, firms usually have less than 10 customer segments. The external population size is the number of Pareto optimal solutions generated by MMSEA. In this case, for number-of-segments ranging from 2 to 10, about 100 solutions are generated for each number-of-segment. The number of generation is set to 2,500 because after 2,500 generations, the improvement of the Pareto front is very small. We tried different crossover rate (0.1 to 0.9 at 0.1 interval) during the evaluation and found that the value of 0.5 gave a good tradeoff of solution quality and convergence speed. The mutation rate changes dynamically and is determined by the number of customers and the distance order of its link edge. If the distance order of link edge is larger, the mutation rate is higher. It makes sense because the chance that two customers should be in different segments is higher if distance between them is bigger. In our evaluation, it works better than a fixed mutation rate.

A summary of generated solutions is shown in Table 5.5. Figure 5.7 shows the results of 4-cluster and 8-cluster segmentation. The K-mean results for 4-cluster and 8-cluster solutions are also displayed for comparison. Even though MMSEA optimizes multiple objectives simultaneously for different number-of-segments, the generated Pareto front is very close to the results of the best of the K-mean runs. This evidence proves the efficiency and effectiveness of MMSEA. It can be seen that MMSEA gives a holistic view of the solution space that covers the extreme optimal solutions for each segmentation criterion.

Number-of-segments	Number of Solutions Generated
2	61
3	107
4	145
5	134
6	128
7	120
8	129
9	92
10	84
Total	1000

Table 5.5 Solution Sizes of MMSEA

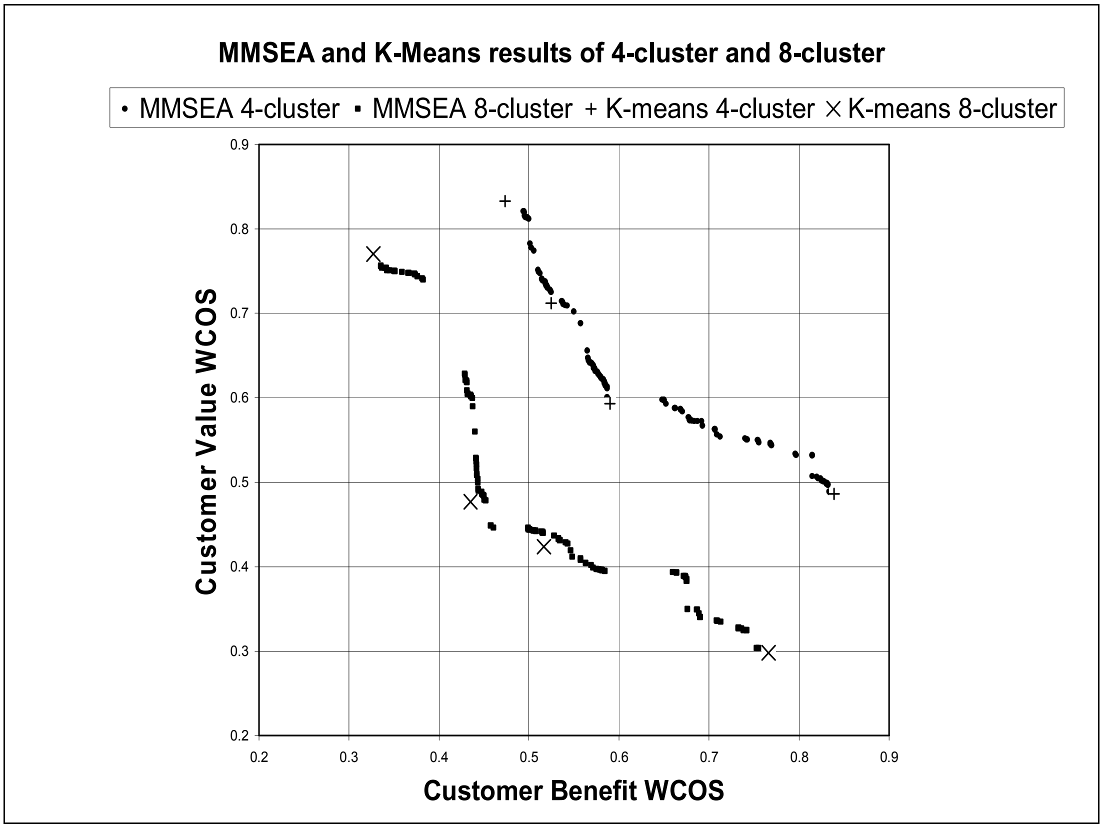


Figure 5.7 4-Segment and 8-Segment Solutions

The results for a total of 1,000 solutions for number-of-segments values ranging from 2 to 10 are shown in Figure 5.8. The top right cluster of points are the Pareto front for the 2-cluster segmentation solutions and the bottom left points are for the 10-cluster segmentation solutions. Unlike most existing multicriterion segmentation methods that require users to make an up-front decision on criteria weights or to provide multiple-to-single criterion transformation functions, MMSEA directly optimizes for multiple criteria and gives a holistic view of the solution space for a range of number-of-segments values. With the aid of a multidimensional visualization tool, it is easy for users to specify constraints on the generated solution set and only investigate the solutions that meet particular requirements for the segmentation problem. Much desired segmentation flexibility is available with this holistic view. Segmentation criterion, segment size, segment/customer level attribute statistics and number-of-segments can be used as decision variables for solution screening. For example, one could specify different weights for different criteria in a weighted sum objective function and choose the corresponding "best" solution. Additionally, the shapes of the Pareto front for a range of number-of-segments values give insight into the tradeoffs for solutions with the same number-of-segments or with different number-of-segments.

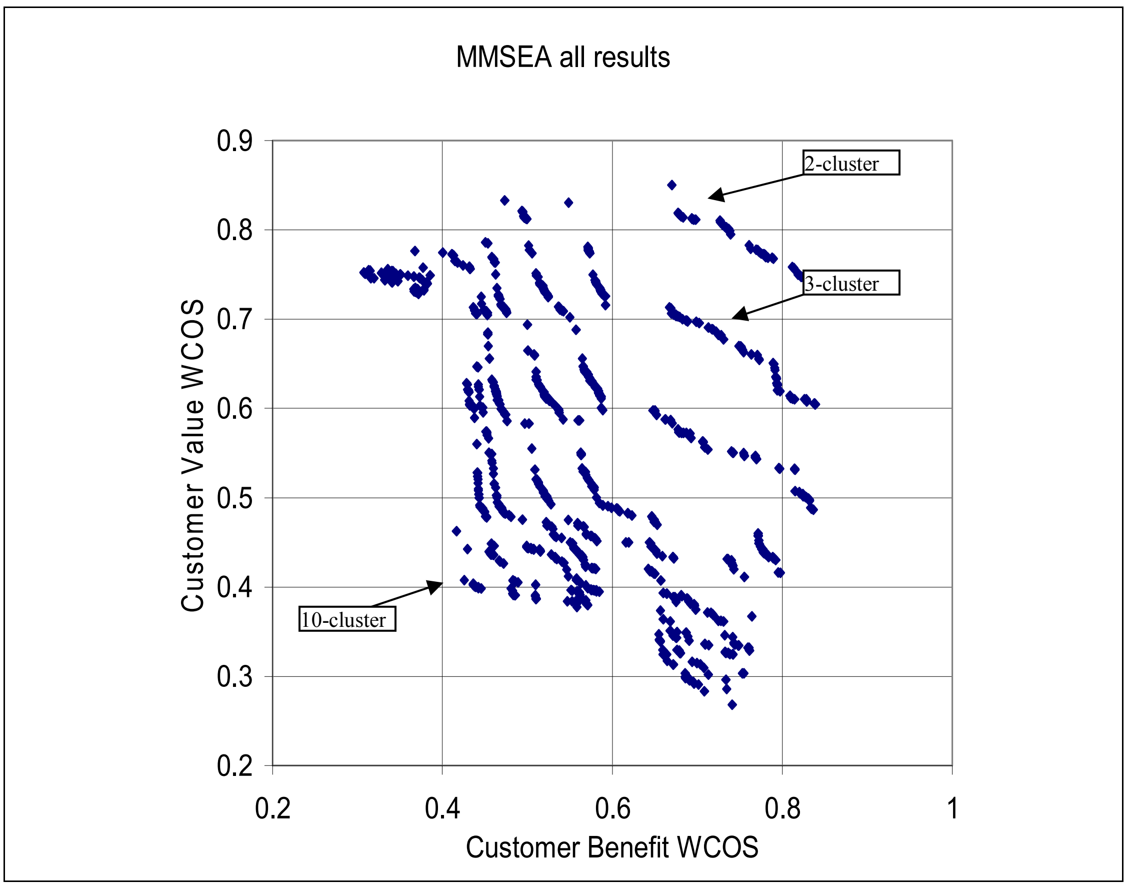


Figure 5.8 All Segmentation Results of MMSEA

5.5 Solution Comparison between K-means and MMSEA

One 4-cluster K-means solution optimized for customer benefit (Table 5.6) was selected to compare with a 4-cluster MMSEA solution (Table 5.7) with "balanced" WCOS values selected from the 4-cluster Pareto front. The values in each cluster column are segment means of the segmentation attributes.

Cluster Number	1	2	3	4
Size	17	15	189	779
Number of churn customers	0	0	6	13
Cluster churn rate	0%	0%	3.2%	1.7%
Percentage of Total	1.7%	1.5%	18.9%	77.9%
Mean monthly minutes of use	2505.3	1346.8	1155.3	317.7
Mean overage minutes of use	491.4	113.6	83.0	17.3
Mean number of outbound voice calls	173.0	60.1	64.5	13.3
Mean number of inbound voice calls	89.5	11.0	20.7	3.8
Mean number of dropped voice call	31.3	16.3	13.8	3.3
Mean number of blocked voice call	16.7	64.7	6.5	2.3
Mean monthly revenue	220.5	121.6	93.1	46.0
Mean total recurring charge	81.5	63.0	63.9	41.9
Months in service	16.4	18.1	17.7	18.6
Mean number of customer care calls	7.7	5.7	4.7	1.2

Table 5.6 K-means 4-cluster Optimized for Customer Benefit WCOS (Customer Benefit WCOS: 0.47, Customer Value WCOS: 0.83)

Cluster Number	1	2	3	4
Size	41	235	258	466
Number of churn customers	0	6	8	5
Cluster churn rate	0%	2.6%	3.1%	0.9%
Percentage of total	4.1%	23.5%	25.8%	46.6%
Mean monthly minutes of use	2146.3	983.2	266.7	302.1
Mean overage minutes of use	334.2	58.7	18.8	14.7
Mean number of outbound voice calls	116.0	53.7	12.8	12.5
Mean number of inbound voice calls	42.4	17.8	3.9	3.5
Mean number of dropped voice call	25.7	11.6	3.2	2.8
Mean number of blocked voice call	25.5	7.1	1.9	2.3
Mean monthly revenue	216.2	80.5	46.4	41.4
Mean total recurring charge	91.2	61.3	39.2	40.3
Months in service	18.2	16.9	30.2	12.7
Mean number of customer care calls	8.1	5.0	0.6	0.7

Table 5.7 MMSEA 4-cluster with a balanced WCOS
(Customer Benefit WCOS: 0.59, Customer Value WCOS: 0.60)

The K-means segmentation solution has two small clusters of heavy users that account for only 3.2% of the total sample customers. Cluster 1 differs from others on its relatively large values on almost all attributes. As expected, the 4 clusters have a higher degree of between-cluster heterogeneity based on customer benefit attributes rather than on customer value attributes. For example, the segment mean of “mean overage minutes of use,” with values of 491.4, 113.6, 83 and 17.7, are significantly different. However, for “mean total recurring charge,” “months in service,” and “mean number of customer care calls” the between-cluster heterogeneity is not as large as desired.

The smallest cluster in MMSEA segmentation is also the heaviest user segment but now it represents 4.1% of the sample customers. It is much bigger than the heaviest user cluster from the K-means segmentation (1.7%). The bigger size of this high-value high-benefit customer cluster means that this is a more profitable and effective segment. Cluster 4 is more stable in terms of cluster churn rate (0.9%) than cluster 3 (with a churn rate of 3.1%) because it has better overall customer benefits than cluster 3. Cluster 3 is a good candidate for targeted retention campaigns because its revenues and number of customer calls are higher than that of cluster 4. An explanation of its high churn rate is that the customer benefit is not high: its mean monthly minutes of use is the smallest while its mean overage minutes of use (18.8) is higher than that of cluster 4 (14.7). An intuitive managerial interpretation is that many customers in this segment have the "incorrect" service plan. All 4 clusters have a similar degree of between-cluster heterogeneity on both customer benefit attributes and customer value attributes. For example, the difference in segment mean for "the mean overage minutes of use" between cluster 3 and 4 is smaller in the MMSEA solution than in the K-means solution. However, this small loss results in a much better between-cluster heterogeneity in "mean total recurring charge," "months in service" and "mean number of customer care calls." With the more balanced cluster size and between-cluster heterogeneity of the MMSEA solution, this solution is more desirable than the K-means solution in general. This analysis shows that many insights can be gleaned from a Pareto optimal set of solutions to a multiobjective problem of segmentation.

5.6 Evaluation Conclusion

This chapter presents completed market segmentation from model definition to comparison with an alternative method. The empirical evaluation has a number of interesting observations:

1) The generated Pareto front is a good one because the solutions generated from single objective optimization algorithms are very close to the Pareto front. Theoretically single objective optimization algorithms are more efficient and more effective in finding good solutions than the multiobjective algorithms because the later need to generate a set of solutions.
2) The Pareto optimal solution set really helps in the sense that it gives the explicit tradeoffs among many conflicting objectives. As shown in this two-objective example, there exist many acceptable solutions that represent a good tradeoff in both criteria.
3) The shape of the Pareto front gives some points that are more interesting than others for their geometric properties in the two dimension solution space. The knees in the Pareto front can be a good start for post-optimization analysis.
4) Because of the link-based chromosome representation, the algorithm allows evolutionary operations on chromosomes with different number of segments. This makes the algorithm effective for generating a large number of Pareto optimal solutions.

5) The set of Pareto optimal solutions gives a lot of flexibility in choosing an appropriate segmentation solution for a specific business domain. The diversity of segment profiles and solution objective values gives many insights that might be missed in single solution techniques.

6 EVALUATION TWO: LINEAR PREDICTIVE MARKET SEGMENTATION

We applied the unified market segmentation again on a real business case. The firm is a large services and solutions retailer that has several hundred stores in the United States and Canada. Let's call it BigRetailer in this dissertation. Several years ago the retailer started a premium customer club program that enables it to associate a customer identifier with each transaction. A rich set of customer purchase data has been accumulated since then. An a priori segmentation of customer based on number of transactions and net sales was used but the researchers in the Customer Relationship Management (CRM) division of BigRetailer believed that they could have a better way to leverage the rich data set of customer, product and market to identify growth opportunities and develop localization strategy. We applied the unified market segmentation on the real customer data and compared the results with the state of the art finite mixture model method.

6.1 Segmentation Model

Because of its direct managerial relevance, profit is a good basis for customer segmentation. Bock and Uncles (Bock and Uncles 2002) observed that profitability is often more cost-effective than other segmentation bases. They proposed that in many cases segmentation by profitability is a good choice for normative segmentation with other forms of heterogeneity being of secondary interest. Using profit as the response variable in a predictive model meets the responsiveness criterion of market segmentation.

On the other side, sociodemographic data helps to answer where the customers are and what their profiles look like. Sociodemographic data is important to fulfill the identifiability and accessibility criteria of market segmentation. Segments formed on the basis of sociodemographic data is easy to identify and to communicate with. Because of its actionable property, sociodemographic data plays an important role in market segmentation. Frank et al ((Frank, Massey et al. 1972) discussed a number of sociodemographic descriptors and reviewed related research. Despite the different behavioral profiles of the various demographic and social segments, the researchers then were unsure about the appropriateness of using sociodemographic data as a basis for market segmentation. Specifically, the research of Frank and Massy (Frank and Massy 1967; McCann 1974) and McCann (McCann 1974) found a weak relationship between demographic data and customer response to many specific marketing mixes. Because above segmentation approaches did not simultaneously optimize on both predictive and descriptive criteria, those approaches can be improved from the new multiobjective perspective proposed in this dissertation.

Sociodemographic data is useful in many scenarios such as direct marketing, but the a priori zip code segments are too coarse for BigRetailer. A number of preliminary studies by BigRetailer researchers and us using census block level data found very weak association between the sociodemographic data and profit or response data. To take full advantage of the sociodemographic data, BigRetailer decided to purchase and use sociodemographic data at the household level.

Consequently, the segmentation model here consists of two types of data: the profit data of each customer and the household level sociodemographic data of the customer.

Two optimization objectives and two constraints are defined for this model. The within cluster omega squared (WCOS) defined in Section 5.2 was used to measure the within segment homogeneity of predictors (the sociodemographic variables). The predictive objective is measured by the residual sum of square of the profit data. To meet the substantiality criterion, a constraint of the minimum size of segment is added. As an instance of the general multicriterion market segmentation problem defined in Section 3.2, we have

z_i = the segment membership of customer i; $i = 1,...,I$, I is the number of customers. There are 1,500 customers in this study.

$z = [z_1, z_2,..., z_I]$ is a vector of decision variables that represent a segmentation solution for all I customers;

(1) $F(z) = [f_1(z), f_2(z)]$ is the objective vector to be optimized;

- $f_1(z)$ = minimize the WCOS of customer sociodemographic basis;

- $f_2(z) =$ minimize the total sum of residual sum of square for all segment level ordinary linear regression models. The response variable is customer profit while the predictor variables are sociodemographic variables;

(2) $G(Z) = [g_1(z), g_2(z)]$ is the constraint vector to be satisfied;

- $g_1(z) =$ the minimum segment size.
- $g_2(z) =$ the number of segments. The user must specify one or a range of number of segments.

(3) A specific segmentation solution $z^* = [z_1^*, z_2^*, ..., z_I^*]$ is Pareto optimal if there does not exist another solution z^l such that

- z^l satisfies (2) and
- $f_j(z^l) \leq f_j(z^*)$ for all $j = 1,2,...,M$ and
- $f_k(z^l) < f_k(z^*)$ for at least one k

(4) The goal is to find one (or more) segmentation solution(s) $z^* = [z_1^*, z_2^*, ..., z_I^*]$ that is (are) Pareto optimal with regard to the objective vector (1) and satisfies constraint vector (2).

Consequently, the segmentation model is a unified market segmentation that optimizes both predictive and descriptive objectives. It meets multiple segmentation criteria from a number of perspectives:

- The responsiveness criterion is satisfied by optimization of response regression model of customer profit.

- The identifiability criterion is satisfied by optimization of within segment homogeneity of sociodemographic predictors.
- The accessibility criterion is satisfied by using sociodemographic variables. The identified segments are easy to access and easy to communicate with.
- The substantiality criterion is satisfied by setting a constraint on the minimum segment size.
- The stability criterion is satisfied by the stability of sociodemographic variables.
- The actionability criterion is a result of segmentation analysis and managerial decisions. It will be discussed in the following sections.

6.2 The Data Set

The data we received is a sample of 43,340 customers from all members of the BigRetailer premium club. The details of the data preprocessing is shown in Appendix A. The customer margin is used as the response. There are seven sociodemographic variables. All variables descriptions are in Table 6.1. We randomly selected 1,500 customers for the empirical evaluation and the descriptive statistics of all variables are listed in Table 6.2.

Role	Variable Name	Description
Predictor	Age	Age of household premier in two year incremental. 17 = Age less than 18. 18 = Age between 18-19, 20 = Age between 20-21, …, 99 = Age > 99.
Predictor	Gender	Gender of household premier. 1 = Female, 0 = Male.
Predictor	Marital	Marital status. 1 = Married, 0 = Single
Predictor	WorkingWoman	Working woman presence in the household. 1 = Yes, 0 = No.
Predictor	Children	Children present in the household. 1 = Yes, 0 = No.
Predictor	Adult	Number of adults in the household.
Predictor	Income	The annual income level of household premier. 1 = less than $15,000 2 = $15,000 - $19,999 3 = $20,000 - $29,999 4 = $30,000 - $39,999 5 = $40,000 - $49,999 6 = $50,000 - $74,999 7 = $75,000 - $99,999 8 = $100,000 - $124,999 9 = GREATER THAN $124,999
Response	Margin	It is the result of transactions occurred approximately from July 2005 to December 2005.

Table 6.1 Model Variable Description

	Minimum	Maximum	Mean	Std. Deviation
Age	18	86	46.80	12.67
Gender	0	1	.72	.45
Marital	0	1	.79	.411
WorkingWoman	0	1	.48	.50
Children	0	1	.57	.50
Adult	1	4	2.17	.71
Income	1	9	6.28	2.05
Margin	0.84	117.22	14.49	11.98

Table 6.2 Descriptive Statistics of Model Variables

The margin variable is transformed using ln() function to reduce its skewness and improve its normality. Because the standardized data performs well in clustering algorithms (Millgan and Cooper 1988), all variables are standardized using z-score for the segmentation process. The collinearity test result is fine. The ordinary linear regression (Table 6.3 and Table 6.4) on the 1500 customers shows a very weak relation between the margin and the sociodemographic predictors. Though some predictors are significant, the goodness of model fit is not good because the seven predictors only explain about 3% of the margin variance. We attribute this weak relationship to the heterogeneity of customer sociodemographic variables. A better prediction model is expected at the segment level.

R Squared	0.031
Adjusted R Squared	0.026
Regression Sum of Square (df = 7)	45.24
Residual Sum of Square (df = 1493)	1435.88
Total Sum of Square (df = 1500)	1481.12

Table 6.3 Model Summary

Variable	Coefficients	Significant
Age	-0.143	0.000
Gender	0.057	0.029
Marital	-0.030	0.292
WorkingWoman	-0.028	0.279
Children	-0.090	0.001
Adult	0.021	0.435
Income	0.086	0.001

Table 6.4 Coefficients of Predictors

6.3 Finite Mixture Model (FMM) Results

The finite mixture model is a class of statistical model methods that assume the data is a mixture of subgroups that have different density function (Dillon and Kumar 1994).

Maximum likelihood methods are usually used to estimate the parameters of finite mixture models. Both expectation-maximization (EM) and numerical optimization are used but EM methods are more popular because of its simplicity (Titterington 1990; Mooijaart and Van Der Heijden 1992). FMM has been used in a number of market segmentation studies and represents the state of the art of market segmentation (Wedel and Kamakura 2000).

There are a number of implementations of finite mixture models ((Wedel and Kamakura 2000). We use the open source implementation developed by Leisch and Grun (Leisch 2004; Grun and Leisch 2006). We ran the algorithm five times and used the best result that has the maximum likelihood of the ordinary linear regression.

Table 6.5 to Table 6.8 show the 3-segment solution results.

Objectives		Segment Sizes		
Deviance	WCOS	1	2	3
796.46	0.9985	1164	172	164

Table 6.5 3-Segment Model Summary

R Squared		0.158
Deviance(Residual Sum of Square)		585.99
Variable	**Coefficients**	**Significant**
Age	-0.155	0.000
Gender	0.105	0.000
Marital	0.144	0.000
WorkingWoman	0.050	0.021
Children	0.143	0.000
Adult	-0.069	0.003
Income	-0.151	0.000

Table 6.6 Segment 1 Results

R Squared		0.636
Deviance(Residual Sum of Square)		170.59
Variable	**Coefficients**	**Significant**
Age	0.298	0.001
Gender	0.202	0.017
Marital	-0.475	0.000
WorkingWoman	0.152	0.071
Children	-0.721	0.000
Adult	0.081	0.340
Income	0.403	0.000

Table 6.7 Segment 2 Results

R Squared		0.874
Deviance(Residual Sum of Square)		39.88
Variable	**Coefficients**	**Significant**
Age	-0.630	0.000
Gender	-0.0243	0.000
Marital	-0.091	0.042
WorkingWoman	-0.534	0.000
Children	-0.306	0.000
Adult	0.151	0.001
Income	0.456	0.000

Table 6.8 Segment 3 Results

The segment level regressions improve a lot compared with the regression of aggregated data. The R squared value increases from 0.031 to 0.874 (segment 3), 0.636 (segment 2) and 0.158 (segment 1). At the segment level, most sociodemographic variables are significant. Nonetheless, because the FMM model only optimizes predictive power, the within segment homogeneity is high. The within cluster omega squared (WCOS) is almost 1.

The results from the 4-segment solution are shown in Table 6.9 to Table 6.13.

Objectives	Segment Sizes				
Deviance	WCOS	1	2	3	4
530.47	0.9974	809	600	60	31

Table 6.9 4-Segment Model Summary

R Squared	0.655	
Deviance(Residual Sum of Square)	343.64	
Variable	Coefficients	Significant
Age	0.060	0.012
Gender	0.063	0.008
Marital	-0.046	0.068
WorkingWoman	0.552	0.000
Children	-0.007	0.789
Adult	-0.066	0.007
Income	-0.062	0.010

Table 6.10 Segment 1 Results

R Squared	0.722	
Deviance(Residual Sum of Square)	146.39	
Variable	Coefficients	Significant
Age	-0.480	0.000
Gender	-0.159	0.000
Marital	-0.017	0.461
WorkingWoman	-0.576	0.000
Children	-0.149	0.000
Adult	0.081	0.000
Income	0.197	0.000

Table 6.11 Segment 2 Results

R Squared		0.835
Deviance(Residual Sum of Square)		38.00
Variable	Coefficients	Significant
Age	0.066	0.672
Gender	0.535	0.000
Marital	-0.211	0.165
WorkingWoman	-0.585	0.071
Children	-0.796	0.000
Adult	0.549	0.000
Income	0.573	0.000

Table 6.12 Segment 3 Results

R Squared		0.976
Deviance(Residual Sum of Square)		2.45
Variable	Coefficients	Significant
Age	-0.185	0.062
Gender	0.443	0.000
Marital	0.360	0.165
WorkingWoman	-0.346	0.001
Children	0.725	0.000
Adult	-0.226	0.002
Income	-0.345	0.000

Table 6.13 Segment 4 Results

4-segment solutions have much better R squared values compared with 3-segment solutions. For the large segments, the linear regression models are very good. Nonetheless, the within segment heterogeneity is still high with a within cluster omega square (WCOS) value close to 1. We will compare the 3-segment and 4-segment solutions of FMM to the corresponding solutions of MMSEA.

6.4 MMSEA Results

6.4.1 The Results of 3-segment to 8-segment Solutions

Traditionally, BigRetailer segments customers from 3-segment to 8-segment. MMSEA generated a set of Pareto optimal solutions for each number-of-clusters value in a single run. The initial solution set is created using a mixture of four methods: clusterwise regression, clusterwise regression followed by k-means, k-means followed by clusterwise regression, and K-means. The algorithm parameters and summary of generated solutions is shown in Table 6.14 and Table 6.15.

Parameter Name	Parameter Value
Number of Generation	10,000
Internal population Size	40
External population Size	1,000
Number-of-clusters	Min: 3 Max: 8
Crossover Rate	0.5
Minimum Segment Size	30

Table 6.14 MMSEA Parameters

Number-of-clusters	Number of Solutions Generated
3	146
4	160
5	184
6	183
7	181
8	146
Total	1000

Table 6.15 Solution Sizes of MMSEA

The results from the 1,000 solutions for number-of-clusters ranging from 3 to 8 are shown in Figure 6.1. The top right cluster of points are the Pareto front for the 3-cluster segmentation solutions and the bottom left points are for the 8-cluster segmentation solutions. The diagram shows the 1000 solutions in the solution space measured by within segment heterogeneity (WCOS -- Within Cluster Omega Squared) and predictive performance (Deviance – the residual sum of square of the ordinary linear regression). The holistic view enables decision makers to look at the profile of each solution, to compare solutions with the same number of segments, and to compare solutions of different number of segments. The tradeoffs are explicated by the characteristics of solutions and the Pareto fronts they formed.

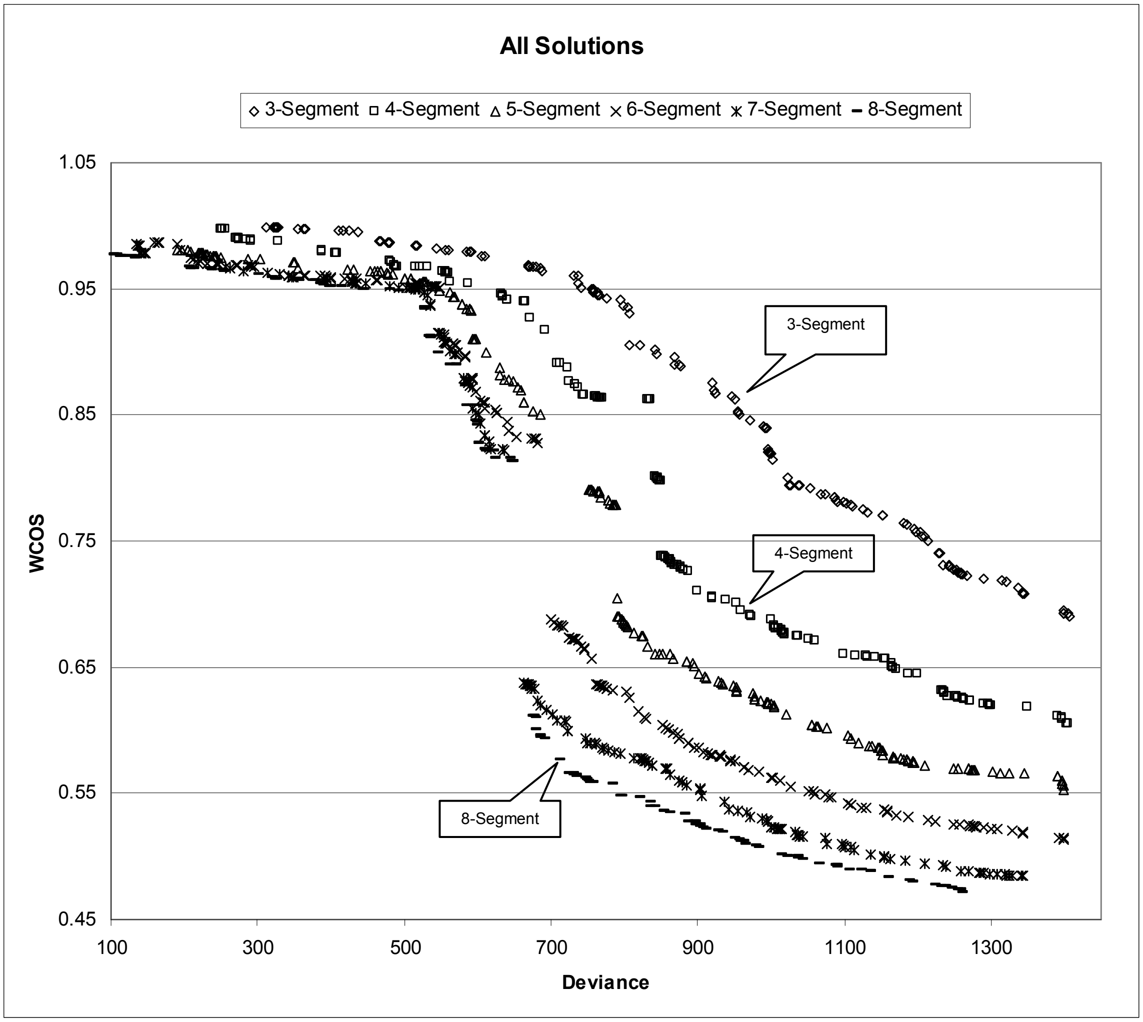


Figure 6.1 All Solutions from 3-segment to 8-segment

6.4.2 3-segment and 4-segment Solutions

Because we want to compare the MMSEA results with FMM results and too many segments could be confusing, we ran the algorithm to generate 3-segment and 4-segment solutions and set the number of generation to 20,000. In reality, decision makers may also want to iteratively run the algorithm to refine their search for best solutions. The effectiveness of MMSEA is shown in Figure 6.2 and Figure 6.3.

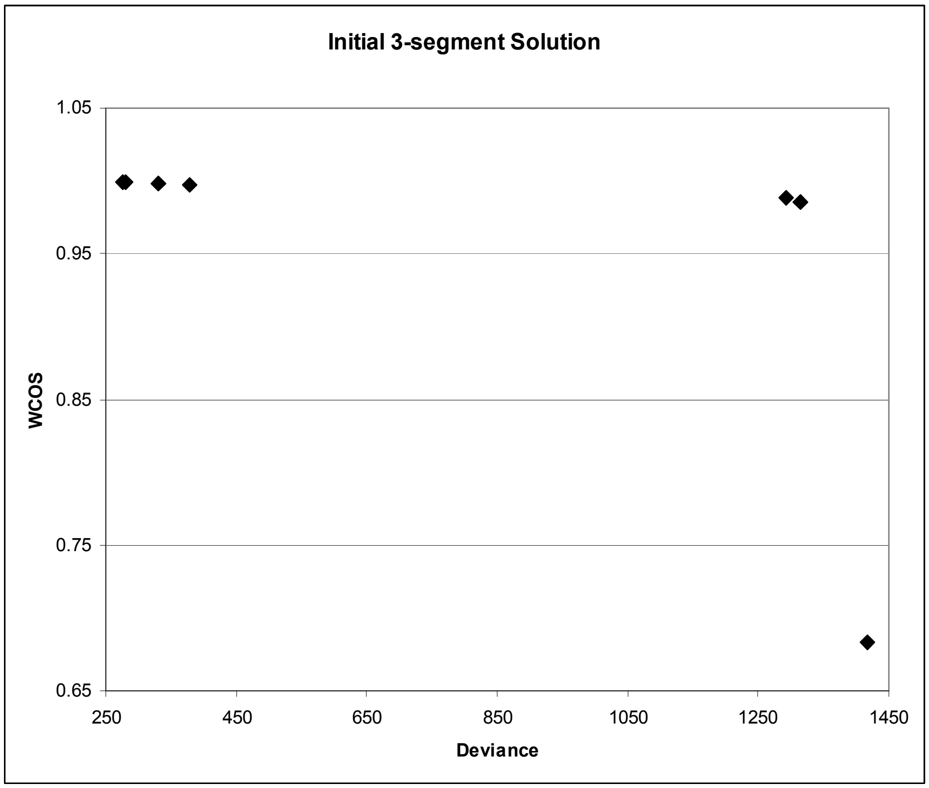


Figure 6.2 Initial 3-segment Solutions

Figure 6.2 shows the initial Pareto optimal solutions. There are only 7 3-segment solutions generated from clusterwise regression, K-means and their mix. Figure 6.3 shows the solutions after 100 generations, 500 generations, 3,000 generations, 10,000 generations and 20,000 generations. The algorithm effectively moves the Pareto front to the bottom left where solutions are desired. Those solutions do not lose too much in one objective while gaining a lot in another objective. Usually they represent good tradeoffs of conflicting objectives and are the solutions decision makers are looking for.

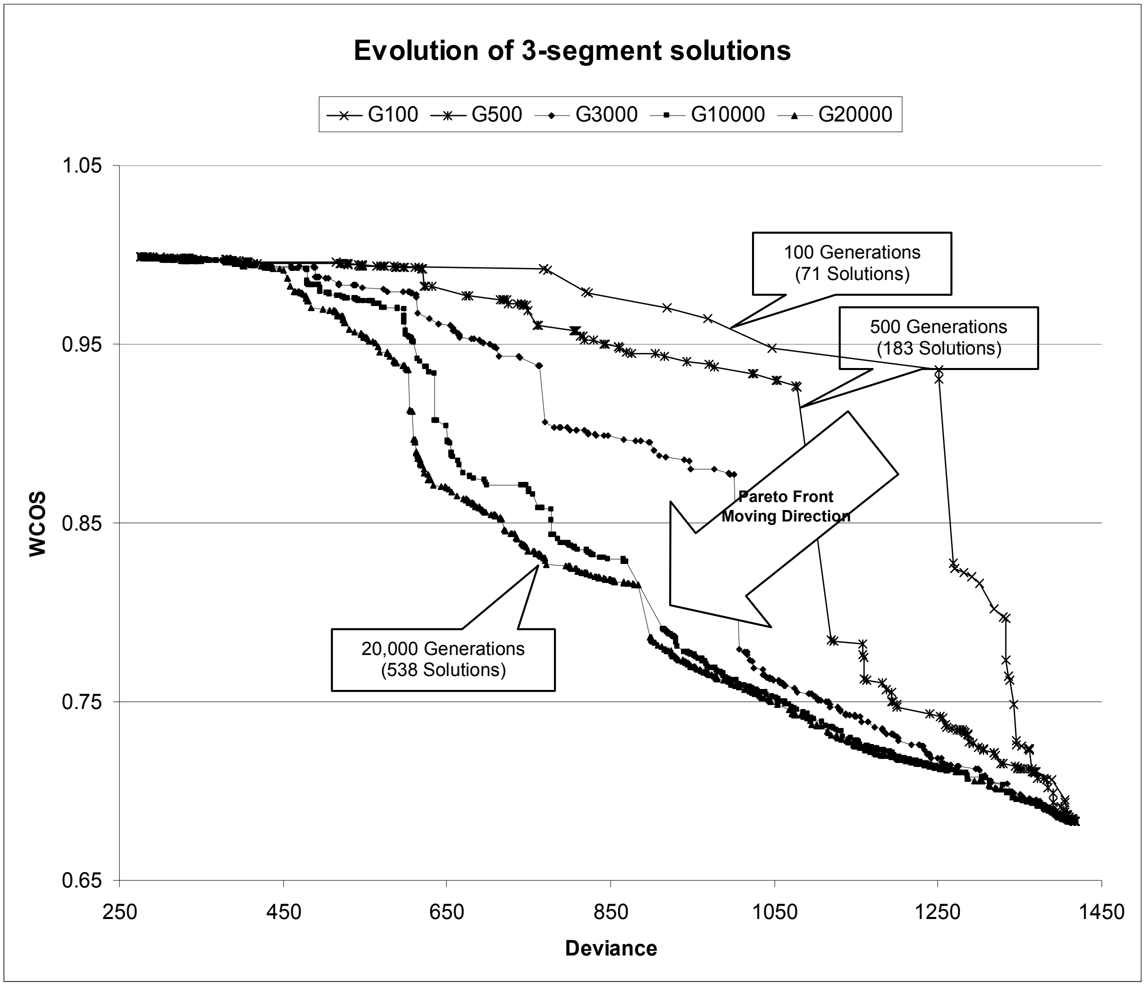


Figure 6.3 Evolution of 3-segment Solutions

Similar results are found in 4-segment solutions (Figure 6.4 and Figure 6.5) and solutions of other numbers of segments.

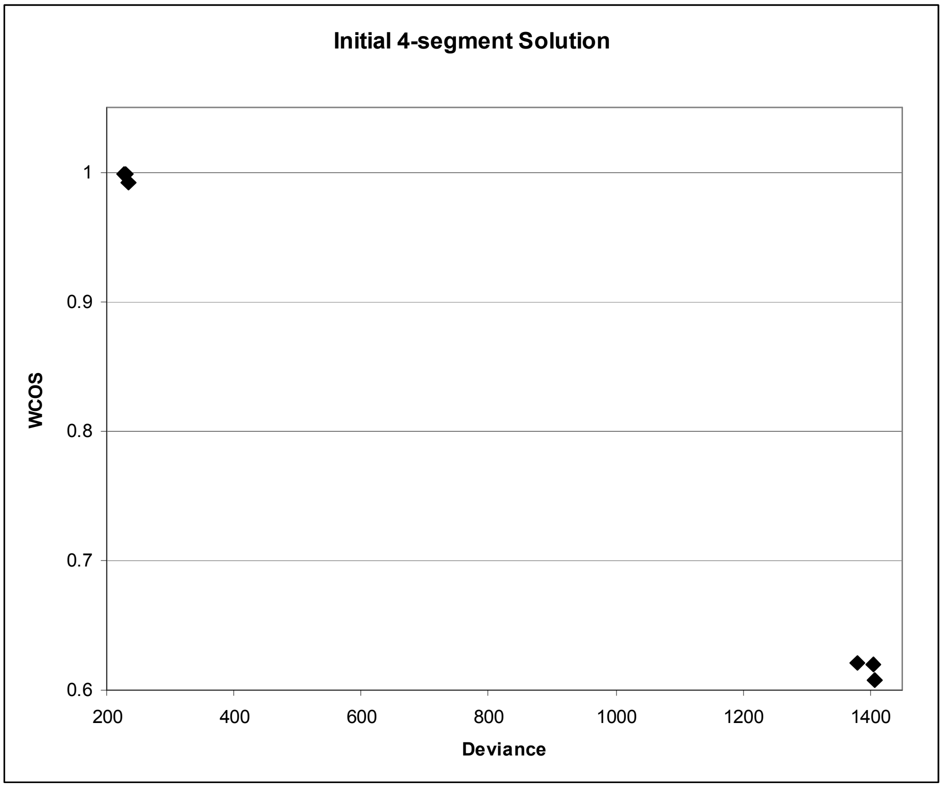


Figure 6.4 Initial 4-Segment Solutions

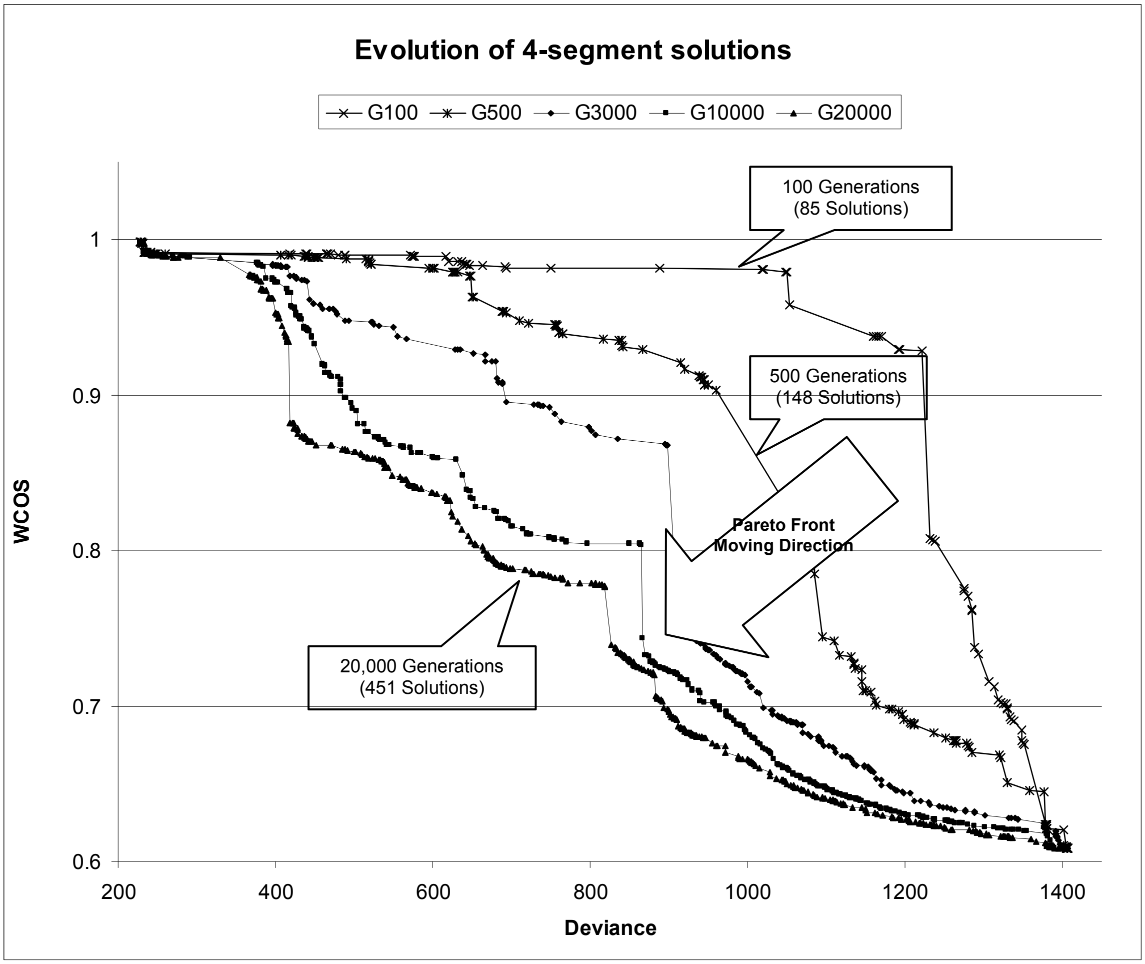


Figure 6.5 Evolution of 4-segment Solutions

In the 4-segment case, after 20,000 generations MMSEA finds 451 Pareto optimal solutions (Figure 6.5) from the original six solutions (Figure 6.4). An observation is that there is no initial solution in the middle of the solution space. The MMSEA is able to close the gap between extreme solutions on the two ends and moves the Pareto front in the desired direction.

Because the search only focuses on 3-segment and 4-segment solutions, it is more efficient than a search of a bigger range of number of segments. Figure 6.6 compares the 3-segment solutions after 10,000 generations for two searches. To make the comparison fair, the only changed parameter is the range of number of segments – even the initial solution set stays unchanged. The 3-segment to 8-segment search generated 146 solutions while the 3-segemnt and 4-segment search generated 501 solutions. Not only do more solutions make the Pareto front smoother, their objective values are much better – especially in the middle part.

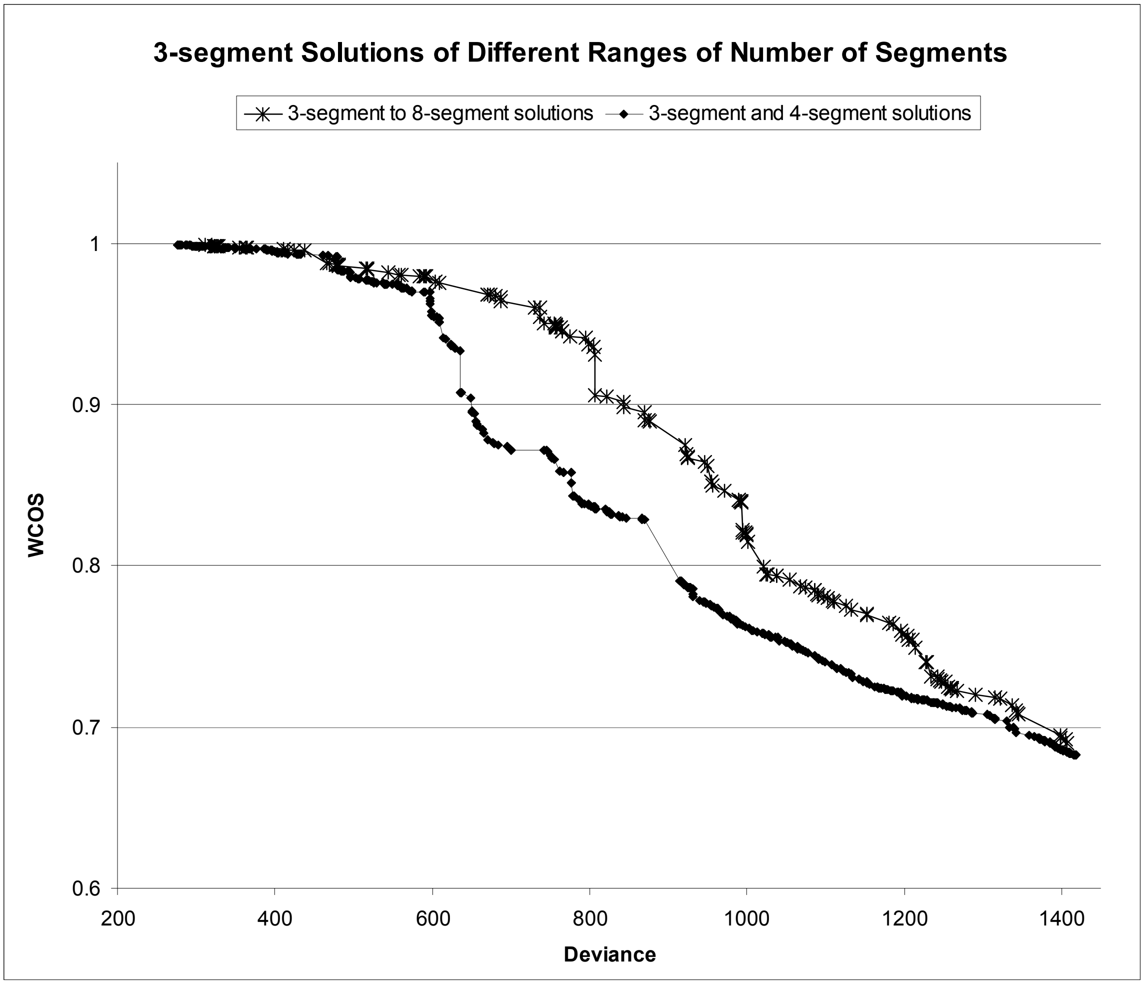


Figure 6.6 Efficiency of Narrowed Search

6.5 Comparison between MMSEA and FMM

The results of MMSEA can be compared with FMM in a number of ways. In the solution space, we compare the Pareto optimal solution set of MMSEA to the single optimal solution found by FMM. If we compare the suggested solutions (the solutions located at the knee of the Pareto front) with the FMM solution, the segment profiles and coefficients of predictors are important measures to decision makers. Only 3-segment and 4-segment solutions are presented here but the conclusion is generally true for the solutions with different number of segments.

6.5.1 Optimization Objectives

Figure 6.7 represents one of the most important results of this research: MMSEA can not only find a set of representative Pareto optimal solutions, but it may also find solutions that are better in both conflicting solutions than the single objective algorithm such as finite mixture model. The set of Pareto optimal solutions is the result of the unified multicriterion definition of the market segmentation problem. The better solutions in both objectives are attributed to the effectiveness and efficiency of multiobjective optimization of the evolutionary algorithm. MMSEA optimizes multiple objectives simultaneously and searches for global optimal solutions.

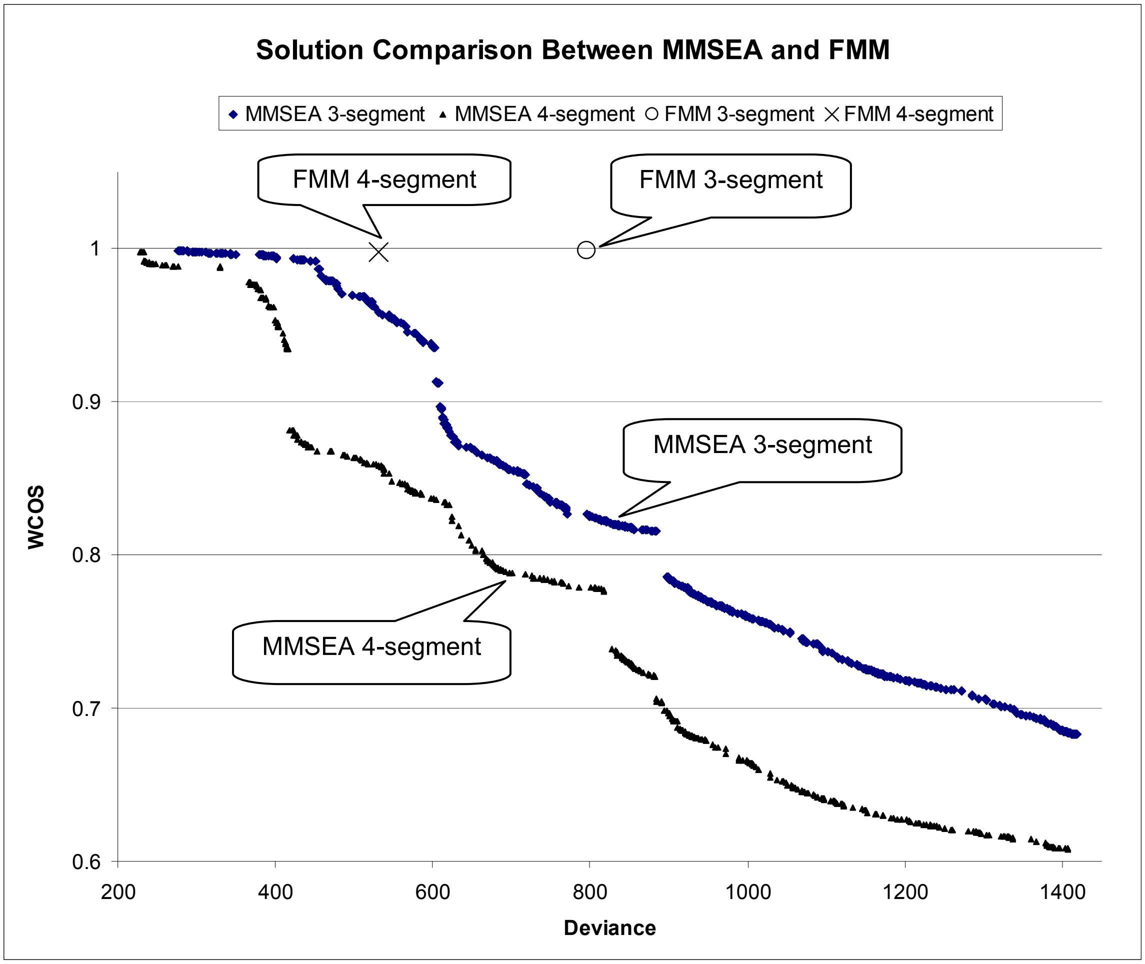


Figure 6.7 Solution Comparison between MMSEA and FMM

Another observation is that, being a single objective optimization method, the finite mixture model approach usually only achieves a good result in one objective. As shown in Figure 6.7, the 3-segment and 4-segment solutions of the finite mixture model have relatively small values in deviance. But the within segment homogeneity, measured by WCOS (within cluster omega squared), is much less than satisfactory.

The Pareto front allows decision makers to choose solutions from many conflicting objectives almost smoothly. There are some gaps in the Pareto front, but within the two extreme solutions of any Pareto front, there are hundreds of solutions to choose from. Each solution may have different segment sizes and segment profiles. The chance is high for the decision makers to find a good solution that meets managerial and institutional constraints.

6.5.2 3-segment Solution Comparison

In this 3-segment solution comparison, we choose a MMSEA solution that has predictive power similar to the FMM solution. The chosen MMSEA solution has a deviance value of 771.14 and a WCOS value of 0.83. The two solutions are compared side by side in Table 6.16. It is difficult to make a fair comparison between the two solutions because they can be compared from many aspects ranging from segment sizes, segment profiles, regression performance, regression coefficients, and coefficient significance. Choosing two segments for comparison is not an easy decision because there are many ways to do so. Nonetheless, we use segment size as the measurement to compare segments from the two solutions. For both solutions, segment 1 has the largest size while segment 3 has the smallest size. Comparing two segments with similar predictive performance or similar response mean also makes sense. We will make such comparisons if it is appropriate.

MMSEA		FMM	
Deviance	771.14	Deviance	796.46
WCOS	0.826	WCOS	0.999
Segment	**Size**	**Segment**	**Size**
1	882	Segment 1	1164
2	355	Segment 2	172
3	263	Segment 3	164

Table 6.16 Comparison of 3-segment Solutions

The deviance of MMSEA is slightly better than that of FMM, but the segment heterogeneity reduces from almost 1 to 0.826. With 882 customers in the biggest segment and 263 customers in the smallest segment, the MMSEA solution is also better in customer assignment. The FMM has a large segment of 1,164 customers and a small segment with only 264 customers. Table 6.17 to Table 6.22 compare segment level regression models and profiles between the MMSEA solution and FMM solution.

	MMSEA Segment		FMM Segment	
Segment Size	882		1164	
R Squared	0.344		0.158	
Variable	**Coefficient**	**Significant**	**Coefficient**	**Significant**
Age	-0.201	0.000	-0.155	0.000
Gender	0.029	0.236	0.105	0.000
Marital	0.003	0.935	0.144	0.000
WorkingWoman	-0.507	0.000	0.050	0.021
Children	-0.125	0.000	0.143	0.000
Adult	0.066	0.016	-0.069	0.003
Income	0.109	0.000	-0.151	0.000

Table 6.17 Segment 1 Regression Model Comparison

	MMSEA Segment		FMM Segment	
Variable	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
Age	46.68	12.03	46.58	12.60
Gender	0.70	0.46	0.73	0.44
Marital	0.96	0.18	0.79	0.41
WorkingWoman	0.46	0.50	0.48	0.50
Children	0.63	0.48	0.57	0.50
Adult	2.34	0.61	2.17	0.70
Income	6.67	1.83	6.30	2.04
Margin	13.87	10.46	13.07	8.31

Table 6.18 Segment 1 Profile Comparison

	MMSEA Segment		FMM Segment	
Segment Size	355		172	
R Squared	0.806		0.636	
Variable	**Coefficient**	**Significant**	**Coefficient**	**Significant**
Age	-0.113	0.000	0.298	0.001
Gender	0.114	0.000	0.202	0.017
Marital	-0.266	0.000	-0.475	0.000
WorkingWoman	1.073	0.000	0.152	0.071
Children	-0.123	0.000	-0.721	0.000
Adult	-0.026	0.434	0.081	0.340
Income	0.036	0.237	0.403	0.000

Table 6.19 Segment 2 Regression Model Comparison

	MMSEA Segment		FMM Segment	
Variable	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
Age	48.59	13.65	46.64	13.12
Gender	0.73	0.45	0.75	0.43
Marital	0.92	0.27	0.76	0.43
WorkingWoman	0.52	0.50	0.54	0.50
Children	0.51	0.50	0.55	0.50
Adult	2.29	0.62	2.10	0.73
Income	6.39	2.02	6.20	1.92
Margin	16.65	16.05	20.09	21.29

Table 6.20 Segment 2 Profile Comparison

	MMSEA Segment		FMM Segment	
Segment Size	263		164	
R Squared	0.050		0.874	
Variable	**Coefficient**	**Significant**	**Coefficient**	**Significant**
Age	-0.142	0.009	-0.630	0.000
Gender	-0.038	0.579	-0.0243	0.000
Marital	0.002	0.973	-0.091	0.042
WorkingWoman	0.064	0.063	-0.534	0.000
Children	-0.121	0.057	-0.306	0.000
Adult	0.071	0.062	0.151	0.001
Income	-0.001	0.053	0.456	0.000

Table 6.21 Segment 3 Regression Model Comparison

	MMSEA Segment		FMM Segment	
Variable	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
Age	44.78	13.13	47.85	12.78
Gender	0.78	0.42	0.63	0.49
Marital	0.00	0.00	0.77	0.42
WorkingWoman	0.53	0.50	0.45	0.50
Children	0.41	0.49	0.55	0.50
Adult	1.40	0.62	2.18	0.75
Income	4.84	2.15	6.21	2.23
Margin	13.67	9.88	18.72	16.95

Table 6.22 Segment 3 Profile Comparison

The MMSEA solution has much better R Squared values in large segments (segment 1 and segment 2). In segment 2, MMSEA has a segment size that is two times of the FMM segment size, but the R squared value is much better (0.806 vs. 0.636). In the largest segment (segment 1), MMSEA improves the prediction performance dramatically from the R Squared value of 0.158 to 0.344. The big gain is without a cost. In the smallest segment (segment 3), MMSEA has a very weak R Squared value of 0.05 while the FMM has the best R Squared value of 0.874. Arguably, an MMSEA solution is more desired because it achieves a good predictive performance in a large portion (82%) of customers with a loss of predictive performance in a small portion (18%) of customers. While FMM has a good predictive performance in the small segments, a large portion (78%) of customers only has an R Squared value of 0.158. Coincidently, segment 1 of both solutions has the lowest margin value among the three segments.

There are many differences in the regression coefficients and predictor significance between the MMSEA solution and FMM solution. Generally, FMM has more significant predictors (at p-value level of 0.01). In segment 3, MMSEA only has one significant predictor (Age) while FMM has 6 significant predictors. In segment 1, MMSEA has 4 significant predictors while FMM has 6 significant predictors. Only in segment 2 does MMSEA have one more significant predictor than FMM. If the p-value level is set to 0.05, then almost all predictors are significant in FMM models. MMSEA only has one significant predictor for segment 1 and five significant predictors for both segment 2 and segment 3. A lower number of predictors not only mean a parsimonious model, it also makes the segment more identifiable.

A post-segmentation regression analysis makes the comparison clearer. We use stepping method to generate the final segment level liner regression model. The entry and removal probability of F is set at 0.05 and 0.10. All variables are standardized thus the regression does not have an intercept term. The results are summarized in Table 6.23.

		MMSEA Solution	FMM Solution
Segment 1	Model	$-.205Age^{**}$ $-.503WorkingWoman^{**}$ $-.122Children^{**}$ $+.109Income^{**}$ $+.065Adult^{*}$	$-.155Age^{**}$ $+.105Gender^{**}$ $+.050WorkingWoman^{*}$ $+.143Children^{**}$ $-.069Adult^{**}$ $-.151Income^{**}$
	R Squared	0.343	0.157
	Margin Mean	163.87	13.07
	Size	882	1164
Segment 2	Model	$-.117Age^{**}$ $+.113Gender^{**}$ $-.263Marital^{**}$ $+1.074WorkingWoman^{**}$ $-.118Children^{**}$	$+.272Age^{**}$ $+.221Gender^{**}$ $-.458Marital^{**}$ $-.778Children^{**}$ $-.431Income^{**}$
	R Squared	0.805	0.627
	Margin Mean	16.05	21.29
	Size	355	172
Segment 3	Model	$-.132Age^{*}$ $-.102Children^{*}$	$-.630Age^{**}$ $-.243Gender^{**}$ $-.091Marital^{*}$ $-.534WorkingWoman^{**}$ $-.306Children^{**}$ $+.151Adult^{**}$ $-.091Income^{**}$
	R Squared	0.034	0.874
	Margin Mean	13.67	18.72
	Size	263	164

Table 6.23 Comparison of Segment Level Regression Models

** Significant at 0.01, * Significant at 0.05

The segment profiles of the MMSEA solution give better identifiability of customers. Segment 1 has average age, highest married ratio, highest working woman rate, highest children presence, highest income and below average margin. Segment 2 has elder age, higher married ratio, less children presence, above average working woman rate, above average income and highest margin. Segment 3 is characterized by unmarried, with fewer adults, less children presence, low income and the least margin. With a positive coefficient value of 1.074, the working woman status of segment 2 is really important. Given the response is about the purchase of a premium product, the result makes sense. The result of the segmentation suggests that we should target segment 2, especially those customers who have a working woman status. We want to investigate segment 1 more because of its big size. A comparison with a 4 or more segment solution might be helpful to check whether more segments are more appropriate or not.

As shown by the WCOS values of MMSEA and FMM, the MMSEA has better within segment homogeneity in each segment. In Table 6.18, Table 6.20, and Table 6.22, the MMSEA has smaller standard deviation values than FMM. Another clue is that for all binary variables (Gender, Marital, WorkingWoman, and Children), the means in FMM segments are similar. One-Way ANOVA tests proved our intuition. For MMSEA (Table 6.24), five predictor means are significantly different among different segments at the 0.01 level while none of the FMM (Table 6.25) predictor means are significantly different among different segments. Even at the 0.05 level, only one FMM predictor mean is significantly different among different segments. This result means that identifiability of MMSEA segmentation is significantly better than that of FMM segmentation.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	2213.645	2	1106.822	6.945	.001
	Within Groups	238587.152	1497	159.377		
	Total	240800.797	1499			
Gender	Between Groups	1.137	2	.568	2.832	.059
	Within Groups	300.381	1497	.201		
	Total	301.517	1499			
Marital	Between Groups	197.175	2	98.588	2649.560	.000
	Within Groups	55.702	1497	.037		
	Total	252.877	1499			
WorkingWoman	Between Groups	1.713	2	.856	3.438	.032
	Within Groups	372.903	1497	.249		
	Total	374.616	1499			
Children	Between Groups	11.724	2	5.862	24.589	.000
	Within Groups	356.874	1497	.238		
	Total	368.597	1499			
Adult	Between Groups	189.682	2	94.841	253.841	.000
	Within Groups	559.315	1497	.374		
	Total	748.997	1499			
Income	Between Groups	677.985	2	338.992	90.657	.000
	Within Groups	5597.729	1497	3.739		
	Total	6275.714	1499			

Table 6.24 One-Way ANOVA of MMSEA Segments

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	204.115	2	102.058	.635	.530
	Within Groups	240596.682	1497	160.719		
	Total	240800.797	1499			
Gender	Between Groups	1.661	2	.830	4.146	.016
	Within Groups	299.857	1497	.200		
	Total	301.517	1499			
Marital	Between Groups	.210	2	.105	.622	.537
	Within Groups	252.667	1497	.169		
	Total	252.877	1499			
WorkingWoman	Between Groups	.746	2	.373	1.493	.225
	Within Groups	373.870	1497	.250		
	Total	374.616	1499			
Children	Between Groups	.100	2	.050	.203	.816
	Within Groups	368.497	1497	.246		
	Total	368.597	1499			
Adult	Between Groups	.733	2	.367	.733	.480
	Within Groups	748.264	1497	.500		
	Total	748.997	1499			
Income	Between Groups	2.564	2	1.282	.306	.736
	Within Groups	6273.150	1497	4.190		
	Total	6275.714	1499			

Table 6.25 One-Way ANOVA of FMM Segments

6.5.3 Suggested 4-segment Solutions

The 3-segment comparison uses a MMSEA solution with predictive power similar to that of the FMM solution. The solution is biased with regard to the predictive performance because it is what FMM solution optimizes for. In this 4-segment comparison, we like to use solutions that have a good tradeoff between the conflicting objectives. Figure 6.8 shows the suggested solutions based on the characteristics of the MMSEA Pareto front.

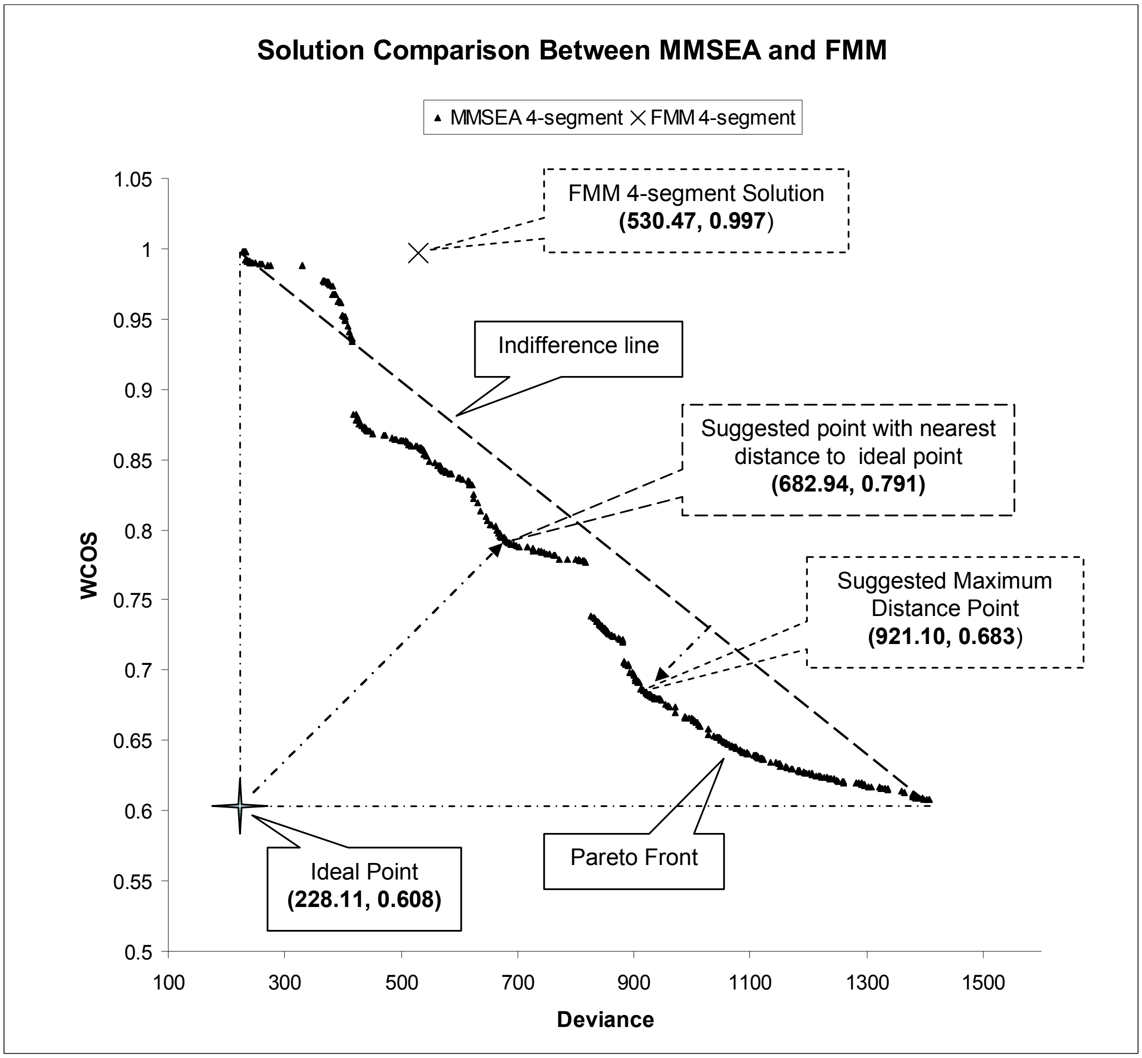


Figure 6.8 MMSEA Suggested 4-segment Solutions

The ideal point at the bottom left shows the unrealistic best solution that has the lowest values of both conflicting objectives. It makes sense that a solution with the nearest distance to this ideal point is an interesting one. The solution with such a property is called ideal knee in this dissertation. The MMSEA solution with a value of (628.94, 0.791) located at the middle of the Pareto front is such an interesting solution that decision makers like to explore.

Nonetheless, there is another type of solutions we like to explore as well. If we assume that the two extreme solutions form an indifference line, i.e., solutions in this line are equally good, solutions sitting below this indifference line are better solutions because they represent a set of solutions that have better gain/loss rate. The solution that has the maximum distance below the indifference line is the one that has the best gain/loss ratio. It is called distance knee in this dissertation. For the suggested distance knee of the Pareto front of 4-segment solutions, the gain and loss values, when compared with FMM solution, are calculated as following:

For the two extreme points of (228.11, 0.998) and (1406.86, 0.608), the range of Deviance is 1178.75 and the range of WCOS is 0.390. Compared with the FMM solution of (530.47, 0.997), the suggested solution (921.10, 0.683) has a worse Deviance value and a better WCOS value. The suggested solution is 33% = (921.10 – 530.47) / 1178.75 worse in Deviance and 81% = (0.997 – 0.683) / 0.390 better in WCOS.

The choice between the two suggested solutions is subjected to managerial, institutional, and resource constraints and is case specific. The two solutions can be a good start for further analysis. In this study we compare both solutions with the FMM solution.

6.5.4 4-segment Solution Comparison

We still use the segment size as the order to compare corresponding solutions. Segment 1 is the largest while segment 4 is the smallest for all solutions. The model summary in Table 6.26 clearly shows the tradeoffs made by MMSEA solutions. If the prediction priority is higher, one may choose the ideal knee solution as a start point. Otherwise, the distance knee gives a solution with high within segment homogeneity.

	Ideal Knee	**Distance Knee**	**FMM Solution**
Optimization Objectives			
Deviance	682.94	921.10	534.47
WCOS	0.791	0.683	0.997
Segment Size			
1	495	506	809
2	406	405	600
3	346	309	60
4	253	280	31

Table 6.26 4-segment Model Summary

Another interesting finding in Table 6.26 is that MMSEA solutions do not have segments with extreme size (too small or too big). This is because segments of similar sizes tend to have a smaller summed WCOS value. As the within segment homogeneity improves, it is usually true that customers are more evenly distributed into each segment. As stated before, MMSEA allows decision makers to specify such preference to influence the generated solutions. For solutions generated by FMM, the segment sizes usually differ significantly. Again, the within segment homogeneity measure of the FMM solution is the worst.

Table 6.27 to Table 6.34 shows the comparisons of 4-segment solutions. FMM has the better segment level predictive performance in all segments but the biggest segment. This is not good for FMM models because the biggest FMM segment represents 54% (809 out of 1,500) of total customers. MMSEA solutions have a pattern in terms of predictive power: larger segments usually have better predictive power while predictors in small segments have weak predictive power. This is reasonable because customers that "do not predict well" are grouped together and could be easily identified. The good news is that those customers represent a small portion of total customers. In the ideal knee solution, 83% of the customers (1,247 out of 1,500) have an R Squared value of 0.429 or better. In the distance knee solution, 61% of the customers (911 out of 1500) have an R Squared value of 0.469 or better. The loss of a little predictive power in small segments is the gain of higher segment identifiability – as shown by the gain loss ratio before.

	MMSEA Segment				**FMM Segment**	
	Ideal Knee		**Distance Knee**			
Segment Size	495		506		809	
R Squared	0.478		0.469		0.448	
Variable	**Coefficient**	**Significant**	**Coefficient**	**Significant**	**Coefficient**	**Significant**
Age	-0.202	0.000	-0.162	0.000	0.060	0.012
Gender	0.202	0.000	0.573	0.000	0.063	0.008
Marital	0.188	0.000	0.023	0.729	-0.046	0.068
WorkingWoman	-0.599	0.000	-0.081	0.024	0.552	0.000
Children	-0.028	0.402	-1.010	0.000	-0.007	0.789
Adult	0.038	0.251	0.074	0.070	-0.066	0.007
Income	0.122	0.000	0.297	0.000	-0.062	0.010

Table 6.27 Segment 1 Regression Model Comparison

	MMSEA Segment				FMM Segment	
	Ideal Knee		Distance Knee			
Variable	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
Age	50.07	12.42	43.47	10.60	46.78	13.07
Gender	0.55	0.50	0.92	0.27	0.72	0.45
Marital	0.94	0.24	0.96	0.20	0.77	0.42
WorkingWoman	0.43	0.50	0.44	0.50	0.50	0.50
Children	0.45	0.50	0.92	0.28	0.55	0.50
Adult	2.35	0.62	2.32	0.59	2.16	0.72
Income	6.43	2.03	6.92	1.67	6.25	2.05
Margin	14.31	10.74	12.01	10.50	14.10	11.21

Table 6.28 Segment 1 Profile Comparison

	MMSEA Segment				FMM Segment	
	Ideal Knee		Distance Knee			
Segment Size	406		405		600	
R Squared	0.429		0.532		0.722	
Variable	**Coefficient**	**Significant**	**Coefficient**	**Significant**	**Coefficient**	**Significant**
Age	-0.225	0.000	-0.207	0.000	-0.480	0.000
Gender	-0.391	0.000	0.714	0.000	-0.159	0.000
Marital	0.053	0.509	0.294	0.006	-0.017	0.461
WorkingWoman	-0.425	0.000	0.014	0.714	-0.576	0.000
Children	-0.311	0.000	0.571	0.000	-0.149	0.000
Adult	0.195	0.000	0.031	0.450	0.081	0.000
Income	0.262	0.000	0.155	0.000	0.197	0.000

Table 6.29 Segment 2 Regression Model Comparison

	MMSEA Segment				FMM Segment	
	Ideal Knee		Distance Knee			
Variable	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Age	42.09	9.93	49.96	13.10	46.83	12.16
Gender	0.89	0.31	0.96	0.20	0.74	0.44
Marital	0.98	0.13	1.00	0.07	0.80	0.40
WorkingWoman	0.46	0.50	0.55	0.50	0.47	0.50
Children	0.87	0.34	0.30	0.46	0.57	0.50
Adult	2.33	0.61	2.37	0.60	2.17	0.69
Income	6.90	1.55	6.29	2.05	6.39	2.02
Margin	12.95	10.00	16.57	13.00	14.44	10.97

Table 6.30 Segment 2 Profile Comparison

	MMSEA Segment				FMM Segment	
	Ideal Knee		Distance Knee			
Segment Size	346		309		60	
R Squared	0.799		0.191		0.835	
Variable	**Coefficient**	**Significant**	**Coefficient**	**Significant**	**Coefficient**	**Significant**
Age	-0.139	0.000	-0.201	0.000	0.066	0.672
Gender	0.123	0.000	0.018	0.612	0.535	0.000
Marital	-0.238	0.000	0.478	0.000	-0.211	0.165
WorkingWoman	1.088	0.000	-0.064	0.177	-0.585	0.071
Children	-0.129	0.000	0.064	0.186	-0.796	0.000
Adult	-0.056	0.120	0.106	0.051	0.549	0.000
Income	0.067	0.037	0.055	0.281	0.573	0.000

Table 6.31 Segment 3 Regression Model Comparison

	MMSEA Segment				FMM Segment	
	Ideal Knee		Distance Knee			
Variable	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
Age	48.02	13.60	49.47	13.41	44.70	12.36
Gender	0.72	0.45	0.02	0.15	0.68	0.47
Marital	0.91	0.29	0.94	0.24	0.85	0.36
WorkingWoman	0.55	0.50	0.41	0.49	0.50	0.50
Children	0.49	0.50	0.49	0.50	0.77	0.43
Adult	2.29	0.61	2.31	0.61	2.12	0.67
Income	6.44	2.01	6.48	1.88	6.02	2.23
Margin	17.12	16.18	14.98	10.98	17.21	21.84

Table 6.32 Segment 3 Profile Comparison

	MMSEA Segment				FMM Segment	
	Ideal Knee		Distance Knee			
Segment Size	253		280		31	
R Squared	0.084		0.073		0.976	
Variable	**Coefficient**	**Significant**	**Coefficient**	**Significant**	**Coefficient**	**Significant**
Age	-0.170	0.002	-0.152	0.005	-0.185	0.062
Gender	-0.120	0.087	-0.158	0.027	0.443	0.000
Marital	-0.028	0.578	-0.092	0.048	0.360	0.165
WorkingWoman	0.064	0.309	0.136	0.036	-0.346	0.001
Children	-0.154	0.007	-0.041	0.466	0.725	0.000
Adult	0.112	0.076	0.064	0.262	-0.226	0.002
Income	-0.011	0.838	0.068	0.185	-0.345	0.000

Table 6.33 Segment 4 Regression Model Comparison

	MMSEA Segment				FMM Segment	
	Ideal Knee		Distance Knee			
Variable	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Age	45.06	13.07	45.31	12.97	50.97	12.07
Gender	0.79	0.41	0.79	0.41	0.65	0.49
Marital	0.00	0.00	0.00	0.00	0.77	0.43
WorkingWoman	0.54	0.50	0.55	0.50	0.45	0.51
Children	0.40	0.49	0.40	0.50	0.52	0.51
Adult	1.37	0.60	1.43	0.66	2.26	0.86
Income	4.80	2.14	4.91	2.20	6.03	2.03
Margin	13.73	9.74	15.44	13.28	20.43	19.88

Table 6.34 Segment 4 Profile Comparison

The regression coefficients of MMSEA also show a pattern: the segment with low predictive power has less significant predictors. This makes sense because less significant predictors usually have less prediction power. However, at the aggregated level, the predictive power is low but the number of significant predictors is large. This suggests that low predictive power is not necessarily attributed to the small number of significant predictors. Consequently we may attribute this model parsimony to the improved within segment homogeneity. Table 6.35 summarizes the ordinary linear regression model using the stepwise regression method.

Segment		Ideal Knee	Distance Knee	FMM Solution
1	Model	$-.188Age^{**}$ $+.196Gender^{**}$ $+.192Marital^{**}$ $-.588WorkingWoman^{**}$ $+.120Income^{**}$	$-.163Age^{**}$ $+.580Gender^{**}$ $-.075WorkingWoman^{**}$ $-.992Children^{**}$ $+.305Income^{**}$	$+.056Age^{**}$ $+.064Gender^{**}$ $+.554WorkingWoman^{**}$ $-.082Adult^{**}$ $-.076Income^{**}$
	R Squared	0.475	0.466	0.446
	Margin Mean	14.31	12.01	14.10
	Size	495	506	809
2	Model	$-.234Age^{**}$ $-.381Gender^{**}$ $-.420WorkingWoman^{**}$ $-.298Children^{**}$ $+.197Adult^{**}$ $+.266Income^{**}$	$-.209Age^{**}$ $+.718Gender^{**}$ $+.303Marital^{**}$ $+.565Children^{**}$ $+.154Income^{**}$	$-.481Age^{**}$ $-.160Gender^{**}$ $-.576WorkingWoman^{**}$ $-.150Children^{**}$ $+.074Adult^{**}$ $+.194Income^{**}$
	R Squared	0.428	0.531	0.722
	Margin Mean	12.95	16.57	14.44
	Size	406	405	600
3	Model	$-.134Age^{**}$ $+.120Gender^{**}$ $-.242Marital^{**}$ $+1.090WorkingWoman^{**}$ $-.121Children^{**}$ $+.067Income^{**}$	$-.214Age^{**}$ $+.450Marital^{**}$	$+.513Gender^{**}$ $-.670WorkingWoman^{**}$ $-.848Children^{**}$ $+.490Adult^{**}$ $+.526Income^{**}$
	R Squared	0.797	0.167	0.828
	Margin Mean	17.12	14.98	17.21
	Size	346	309	60
4	Model	$-.167Age^{**}$ $-.178Children^{**}$ $+0.083Adult^{*}$	$-.148Age^{**}$	$+.513Gender^{**}$ $+.361Marital^{**}$ $-.288WorkingWoman^{**}$ $+.837Children^{**}$ $-.263Adult^{**}$ $-.305Income^{**}$
	R Squared	0.073	0.026	0.972
	Margin Mean	13.73	15.44	20.43
	Size	253	280	31

Table 6.35 Comparison of Segment Level Regression Models

** Significant at 0.01, * Significant at 0.05

Segment 3 of the ideal knee solution represents a good target for a future marketing campaign for a number of reasons: it has the highest margin and the highest predictive power. Compared with the population profile, the segment 3 customers are characterized by elder age, more working women, higher married rate, and lightly above average income. The most significant predictor in segment 3 is working woman status that is positively related to margin. Marital status, age, gender, and children presence are additional significantly important predictors.

For similar reasons, segment 2 of the distance knee solution can be a good target. It has the highest margin mean and predictive power. It is more identifiable than the above ideal knee segment. Customers in segment 2 can be described as elder, most are women, (almost all) married, more likely a working woman and less likely with children presence in house.

The choice between the ideal knee and distance knee is really a tradeoff between predictability and identifiability. No matter how powerfully a segment can predictive its margin, if it is not adequately identifiable, the segmentation is less useful. The reverse is also true. The decision is really problem specific and is closely associated with the cost benefit analysis under the managerial and resource constraints. One of the biggest advantages of MMSEA approach is that it enables this analysis naturally.

Again, one-way ANOVA analysis (Table 6.36, Table 6.37, and Table 6.38) shows that the segments of the two MMSEA solutions are significantly different from each other. As expected, distance knee is more identifiable because its WCOS value is the smallest among the three solutions. In this case, the FMM solution's lack of identifiability severely limited its usefulness in practice.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	17836.718	3	5945.573	39.892	.000
	Within Groups	222964.079	1496	149.040		
	Total	240800.797	1499			
Gender	Between Groups	28.440	3	9.480	51.934	.000
	Within Groups	273.077	1496	.183		
	Total	301.517	1499			
Marital	Between Groups	188.717	3	62.906	1466.742	.000
	Within Groups	64.160	1496	.043		
	Total	252.877	1499			
WorkingWoman	Between Groups	4.274	3	1.425	5.755	.001
	Within Groups	370.342	1496	.248		
	Total	374.616	1499			
Children	Between Groups	53.653	3	17.884	84.951	.000
	Within Groups	314.945	1496	.211		
	Total	368.597	1499			
Adult	Between Groups	192.571	3	64.190	172.581	.000
	Within Groups	556.426	1496	.372		
	Total	748.997	1499			
Income	Between Groups	729.975	3	243.325	65.639	.000
	Within Groups	5545.739	1496	3.707		
	Total	6275.714	1499			
Margin	Between Groups	3516.135	3	1172.045	8.282	.000
	Within Groups	211715.290	1496	141.521		
	Total	215231.425	1499			

Table 6.36 One-Way ANOVA of Ideal Knee Solution

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	12454.888	3	4151.629	27.199	.000
	Within Groups	228345.910	1496	152.638		
	Total	240800.797	1499			
Gender	Between Groups	195.253	3	65.084	916.258	.000
	Within Groups	106.265	1496	.071		
	Total	301.517	1499			
Marital	Between Groups	212.892	3	70.964	2655.049	.000
	Within Groups	39.985	1496	.027		
	Total	252.877	1499			
WorkingWoman	Between Groups	4.999	3	1.666	6.745	.000
	Within Groups	369.617	1496	.247		
	Total	374.616	1499			
Children	Between Groups	100.653	3	33.551	187.324	.000
	Within Groups	267.944	1496	.179		
	Total	368.597	1499			
Adult	Between Groups	187.767	3	62.589	166.836	.000
	Within Groups	561.230	1496	.375		
	Total	748.997	1499			
Income	Between Groups	742.164	3	247.388	66.882	.000
	Within Groups	5533.550	1496	3.699		
	Total	6275.714	1499			
Margin	Between Groups	5175.564	3	1725.188	12.287	.000
	Within Groups	210055.861	1496	140.412		
	Total	215231.425	1499			

Table 6.37 One-Way ANOVA of Distance Knee Solution

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	804.065	3	268.022	1.671	.171
	Within Groups	239996.732	1496	160.426		
	Total	240800.797	1499			
Gender	Between Groups	.394	3	.131	.653	.581
	Within Groups	301.123	1496	.201		
	Total	301.517	1499			
Marital	Between Groups	.515	3	.172	1.018	.384
	Within Groups	252.362	1496	.169		
	Total	252.877	1499			
WorkingWoman	Between Groups	.426	3	.142	.568	.636
	Within Groups	374.190	1496	.250		
	Total	374.616	1499			
Children	Between Groups	2.837	3	.946	3.868	.009
	Within Groups	365.760	1496	.244		
	Total	368.597	1499			
Adult	Between Groups	.431	3	.144	.287	.835
	Within Groups	748.566	1496	.500		
	Total	748.997	1499			
Income	Between Groups	11.369	3	3.790	.905	.438
	Within Groups	6264.345	1496	4.187		
	Total	6275.714	1499			
Margin	Between Groups	1660.404	3	553.468	3.877	.009
	Within Groups	213571.021	1496	142.761		
	Total	215231.425	1499			

Table 6.38 One-Way ANOVA of FMM Solution

6.6 Evaluation Conclusion

This chapter presents a detailed evaluation of the multicriterion market segmentation using MMSEA in a real business setting. MMSEA is able to simultaneously optimize two objectives that represent the responsiveness and identifiability criteria of market segmentation. By setting a minimum segment size, the segment solutions are guaranteed to meet the substantiality criterion. Because customers profiled by sociodemographic variables are easy to access and relatively stable in sociodemographic status, the MMSEA segmentation solutions allow decision makers to formulate more effective and efficient market programs. The evaluation of this chapter shows the superiority of the unified multicriterion market segmentation in a number of ways.

First, the unified method is able to give a holistic view of possible solutions that represent different tradeoffs in an explicit way. MMSEA is able to generate a set of solutions with different number-of-segments in a single run. The results represent good diversity in the solution space.

Second, the unified method promotes iterative segmentation analysis. The first iteration with loose parameters and constraints gives a holistic view of the characteristics of possible solutions in the context of a number of conflicting objectives. The narrowed search in later iteration gives more and better solutions with the specified properties. Those solutions not only represent good diversity and better tradeoffs (because of the better objective values in the middle part of the Pareto front), they also allow decision makers to choose solutions among conflicting objectives smoothly.

Third, the combination of the multiobjective optimization problem definition with an effective global search algorithm often results in superior solutions with regard to multiple objectives. In both 3-segment and 4-segment evaluation, the MMSEA is able to find solutions that are better in both objectives than the FMM method.

Fourth, MMSEA allows very flexible constraints such as segment size or range of number of segments. FMM is less flexible in parameter setting.

The evolutionary moving of Pareto front shows that MMSEA is effective and efficient for generating a good number of solutions even from a small number of initial solutions. Especially in the 4-segment evaluation case, MMSEA is able to discover the Pareto front even when there are only six extreme solutions. It fills the gap quickly even after several hundred generations.

Compared with FMM, MMSEA has several disadvantages. At the model level, FMM allows statistic inference, which the MMSEA only partially supports. FMM is a relatively independent algorithm by itself. MMSEA relies on good quality initial solutions generated by other single objective optimization methods. In this evaluation, clusterwise regression and K-means are used to generate the initial solution set. Additionally, MMSEA is a more complex search algorithm and takes longer to run. It is not easy to compare the performance between FMM and MMSEA because FMM stops at local optimal solutions and often finishes in several hundred iterations. MMSEA is a global search algorithm for a NP-hard problem – this means for a non-trivial problem, it may run forever. However, based on the shape of the Pareto front, we found MMSEA makes little progress after 20,000 generations (each generation creates 40 candidate solutions). On a computer with 3.2GHz Pentium D CPU and 2GB memory, the average running time is shown in Table 6.39.

FMM (5 runs)	Each MMSEA Generation	Clusterwise Regression	K-means
21.3 seconds	17.7 seconds (40 candidates)	1.8 hour	< 1 second

Table 6.39 FMM and MMSEA Performance

The total MMSEA running time is calculated as: *Number of clusterwise initial solutions * 1.8 hour + Number of k-means initial solutions * 1 second + number of generations * 188 second.* The clock time used to run 20,000 generations of MMSEA, with 40 initial solutions (30 of them involved clusterwise regression), is about 6 days, 8 hours, and 32 minutes.

7 EVALUATION THREE: LOGISTIC PREDICTIVE MARKET SEGMENTATION

With a rich set of customer purchase data and household level sociodemographic data, BigRetailer wants to improve its promotion responsiveness. Specifically, BigRetailer mailed a discount coupon for a premium product to its premium club members. A response flag is recorded if the customer buys the premium product using the coupon and buys at least one more time without the coupon. A logistic regression model is a common tool for the dichotomous response variable.

7.1 Segmentation Model

Two optimization objectives and two constraints are defined for this model. The within cluster omega squared (WCOS) defined in Section 5.2 was used to measure the segment homogeneity of the predictors (the sociodemographic variables and the purchase data). The responsiveness is measured by the residual deviance of the logistic regression. To meet the substantiality criterion, a constraint of the minimum size of segment is defined. As an instance of the general multicriterion market segmentation problem defined in Section 3.2, we have

z_i = the segment membership of customer i; $i = 1,...,I$, I is the number of customers. There are 1500 customers in this study.

$z = [z_1, z_2, ..., z_I]$ is a vector of decision variables that represent a segmentation solution for all I customers;

(1) $F(z) = [f_1(z), f_2(z)]$ is the objective vector to be optimized;

- $f_1(z)$ = minimize the WCOS of customer sociodemographic basis;
- $f_2(z)$ = minimize the total deviance for all segment level logistic regression models. The response variable is the response flag while the predictor variables are sociodemographic and purchase variables;

(2) $G(Z) = [g_1(z), g_2(z)]$ is the constraint vector to be satisfied;

- $g_1(z)$ = the minimum segment size.
- $g_2(z)$ = the number of segments. The user must specify one or a range of number of segments

(3) A specific segmentation solution $z^* = [z_1^*, z_2^*, ..., z_I^*]$ is Pareto optimal if there does not exist another solution z^l such that

- z^l satisfies (2) and
- $f_j(z^l) \leq f_j(z^*)$ for all $j = 1,2,...,M$ and
- $f_k(z^l) < f_k(z^*)$ for at least one k

(4) The goal is to find one (or more) segmentation solution(s) $z^* = [z_1^*, z_2^*, ..., z_I^*]$ that is (are) Pareto optimal with regard to the objective vector (1) and satisfies constraint vector (2).

Consequently, the segmentation model is a unified market segmentation that optimizes both predictive and descriptive objectives. This unified market segmentation model addresses the multicriterion requirements from a number of perspectives:

- The responsiveness criterion is satisfied by optimization of the logistic regression model in which customer response flag is the dependent variable.
- The identifiability criterion is satisfied by optimization of within segment homogeneity of sociodemographic and purchase predictors.
- The accessibility criterion is satisfied by using sociodemographic variables.
- The substantiality criterion is satisfied by setting a constraint on the minimum segment size.
- The stability of the model depends on the stability of sociodemographic and purchase variables.
- The actionability criterion is a result of segmentation analysis and managerial decisions. It will be discussed in the following sections.

7.2 The Data Set

The data is a sample of 43,340 customers from all members of the BigRetailer premium club. The detail preprocess of data for the logistic regression is shown in Appendix B. The response (dependent) variable is the target flag of the customer. There are three sociodemographic variables and three purchase variable in the final model. All variable descriptions are in Table 7.1. We randomly selected 1500 customers for the empirical evaluation and the descriptive statistics of all variables are listed in Table 7.2. All predictor variables are standardized using z-score for the segmentation process.

Role	Variable Name	Description
Predictor	Age	Age of household premier in two year incremental. 17 = Age less than 18. 18 = Age between 18-19, 20 = Age between 20-21, …, 99 = Age > 99.
Predictor	Children	Children present in the household. 1 = Yes, 0 = No.
Predictor	Income	The annual income level of household premier. 1 = less than \$15,000 2 = \$15,000 - \$19,999 3 = \$20,000 - \$29,999 4 = \$30,000 - \$39,999 5 = \$40,000 - \$49,999 6 = \$50,000 - \$74,999 7 = \$75,000 - \$99,999 8 = \$100,000 - \$124,999 9 = GREATER THAN \$124,999
Predictor	DSLP	Days since last purchase
Predictor	Margin	Customer margin in 6 months
Predictor	Purchase	Number of purchases in 6 months
Response	Target	1 = customer responded to the campaign 0 = customer did not respond to the campaign

Table 7.1 Statistics of Segmentation Variables

	Minimum	Maximum	Mean	Std. Deviation
Age	18	88	46.39	12.61
Children	0	1	0.59	0.493
Income	1	9	6.35	2.059
DSLP	0	184	48.42	47.01
Margin	-9.20	210.89	15.17	14.85
Purchase	0	44	4.94	4.41
Target	0	1	0.62	0.49

Table 7.2 Descriptive Statistics of Model Variables

Table 7.3 and Table 7.4 show the logistic regression results of all 1500 customers.

Cox & Snell R Square	0.212
Nagelkerke R Square	0.288
Percentage Correct	72.7

Table 7.3 Model Summary

Variable	Coefficients	Significant
Age	0.180	0.009
Children	-0.130	0.040
Income	0.025	0.694
DSLP	-0.512	0.000
Margin	-0.141	0.013
Purchase	1.166	0.000
Constant	.739	0.000

Table 7.4 Coefficients of Predictors

7.3 MMSEA Results

We run the MMSEA algorithm to generate solutions from 3-segment to 5-segment. The initial solution set is created using a mixture of four methods: clusterwise logistic regression, clusterwise logistic regression followed by K-means, K-means followed by clusterwise logistic regression, and K-means. The MMSEA algorithm parameters and summary of generation solutions are shown in Table 7.5 and Table 7.6.

Parameter Name	Parameter Value
Number of Generation	20,000
Internal population Size	40
External population Size	1,000
Number-of-clusters	Min: 3 Max: 5
Crossover Rate	0.5
Minimum Segment Size	30

Table 7.5 MMSEA Parameters

Number-of-clusters	Number of Solutions Generated
3	383
4	287
5	321
Total	991

Table 7.6 Solution Sizes of MMSEA

Figure 7.1 shows the 8 initial 3-segment solutions generated by clusterwise logistic regression and K-means in the solutions spaces. Five solutions are in the top left and 2 are in the bottom right. There is only one intermediate result located in the upper right of the solution space. It is not a good solution because its values are high in both objectives. Figure 7.2 shows that the evolution of 3-segment solutions from generation 2,000 to generation 20,000. The MMSEA is able to discover a decent Pareto front from a few initial solutions after 2,000 generations. After 10,000, the scale of improvement is decreased because the solution set is approaching the real Pareto front. Similar results are found for 4-segment solutions (Figure 7.3 and Figure 7.4) and 5-segment solutions (Figure 7.5 and Figure 7.6). Figure 7.7 shows all solution results after 20,000 generations. As before, each Pareto front gives a holistic view of the possible solutions of a specified number-of-segments. Decision makers are able to compare solutions with the same number-of-segments and different number of segments. The tradeoffs of solutions in the solution space are explicated by the characteristics of the Pareto front.

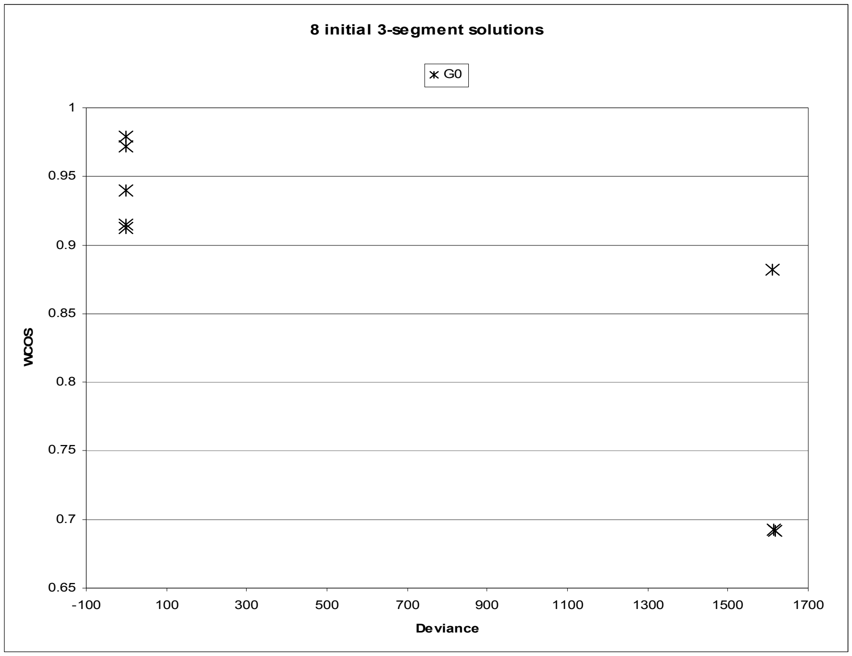


Figure 7.1 Initial 3-segment solutions

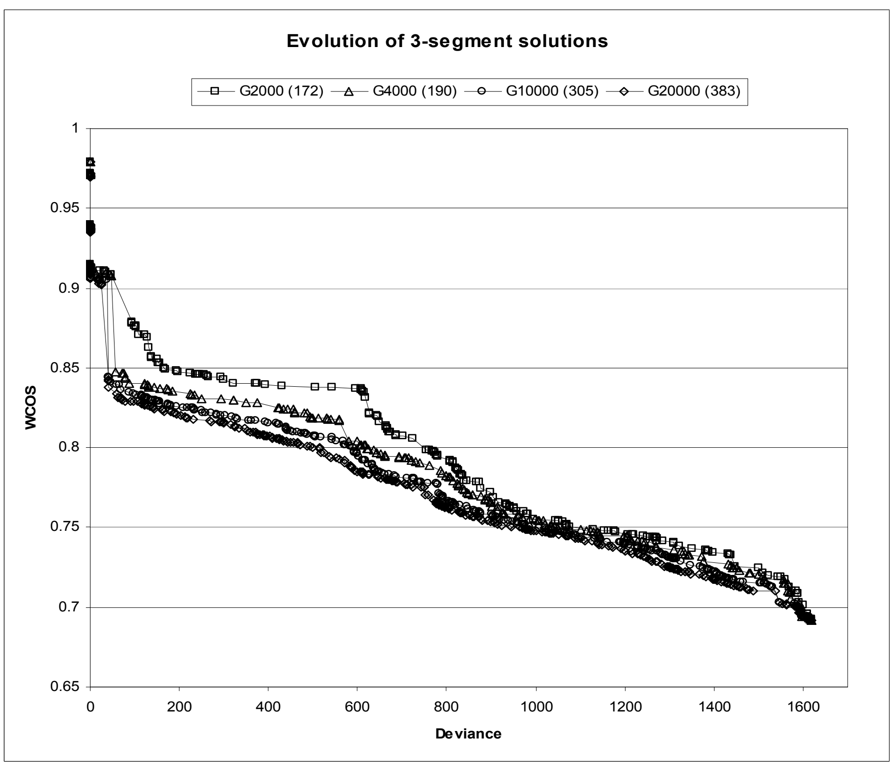


Figure 7.2 Evolution of 3-segment solutions

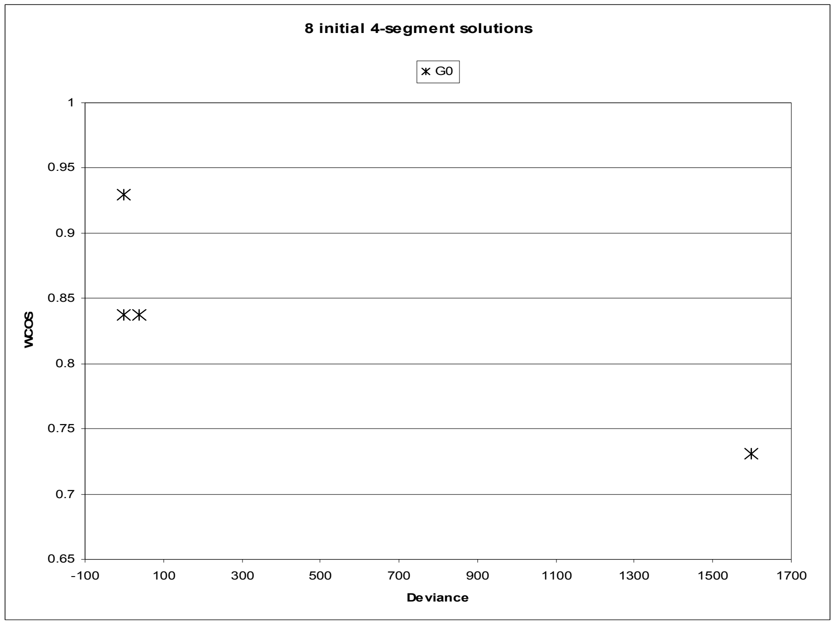


Figure 7.3 8 inital 4-segment solutions

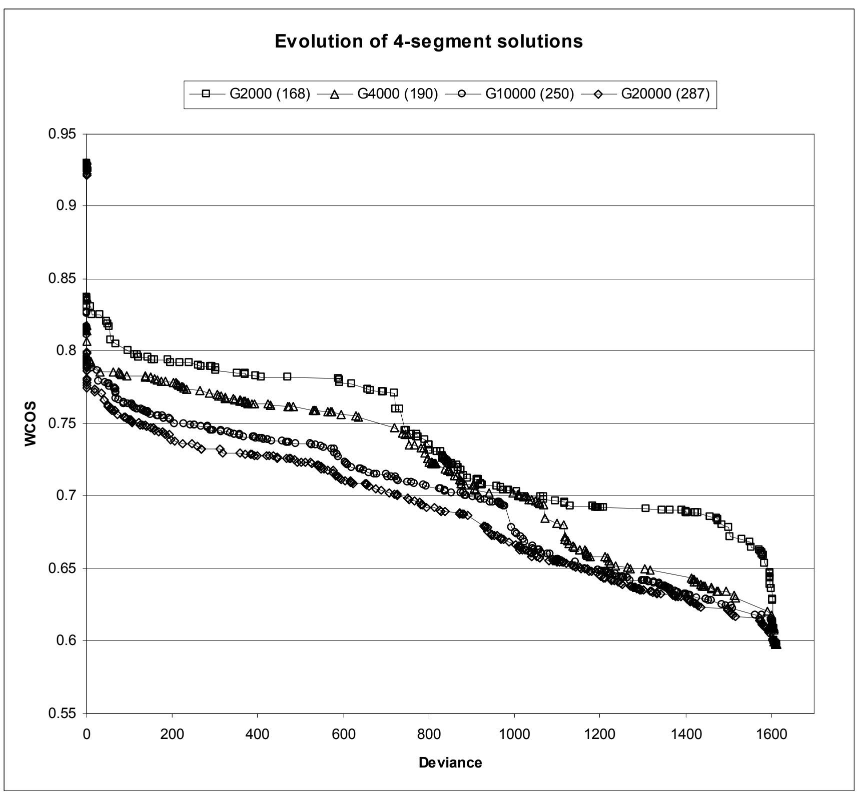


Figure 7.4 Evolution of 4-segment solutions

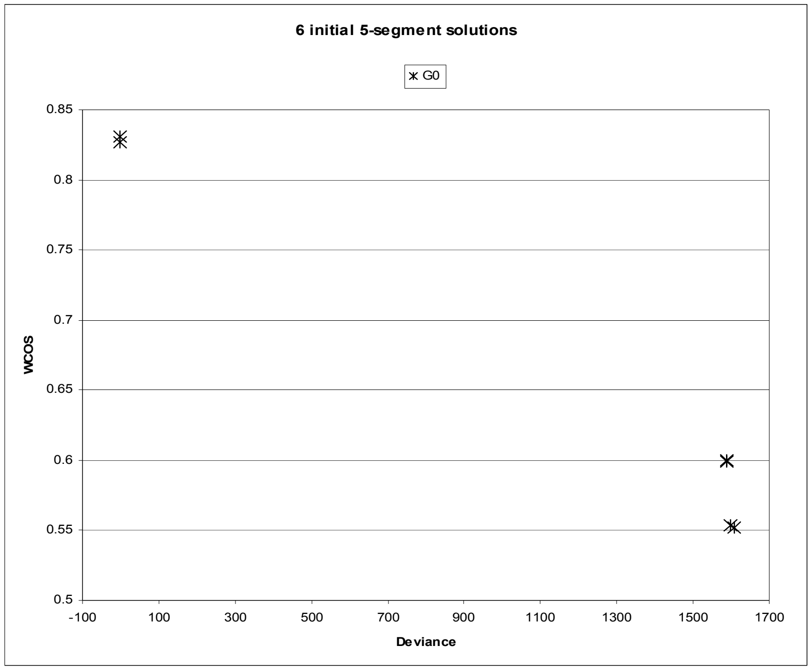


Figure 7.5 6 initial 4-segment solutions

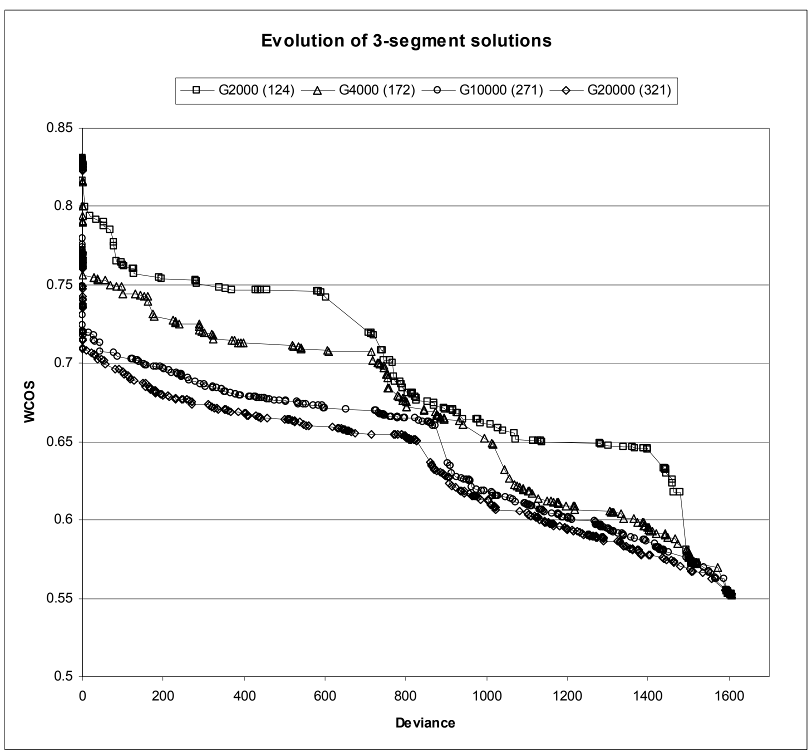


Figure 7.6 Evolution of 5-segment solutions

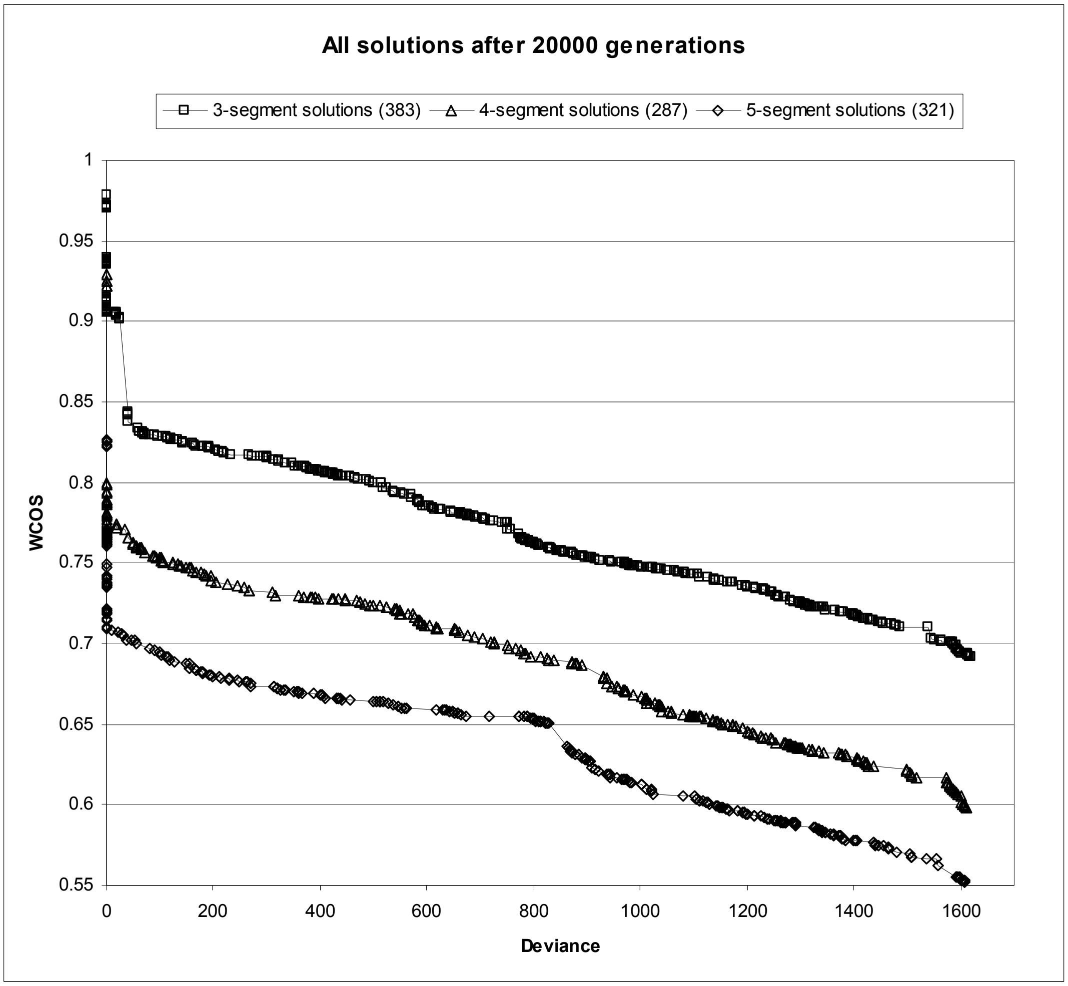


Figure 7.7 All solutions

7.4 Comparison between MMSEA and FMM

In this section we compare the 3-segment and 4-segment solutions between MMSEA and FMM.

7.4.1 Optimization Objectives

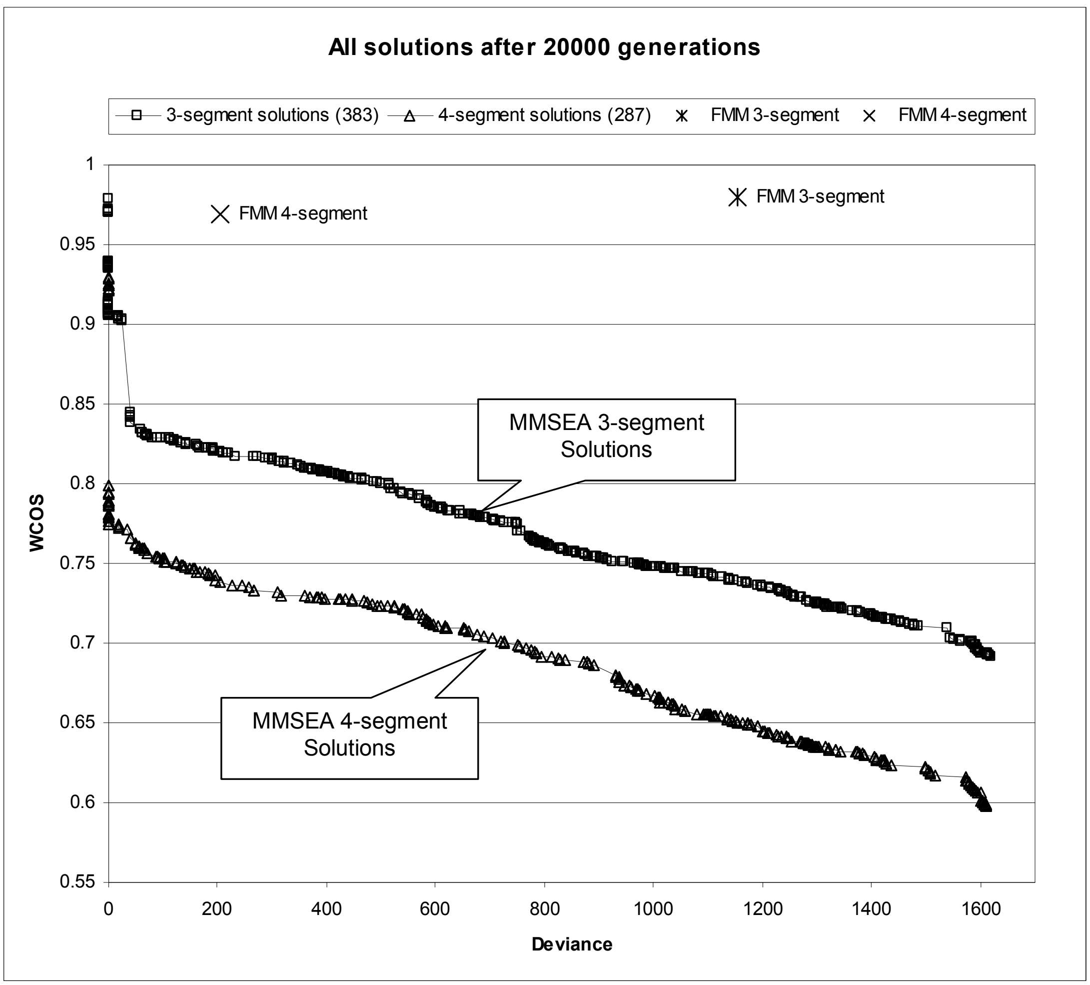


Figure 7.8 Solution Comparison of Logistic Regression

The MMSEA is able to find solutions that are better in both objectives (Figure 7.8).

7.4.2 Comparison of 3-segment Solutions

We select the ideal knee of the MMSEA 3-segment Pareto front. Figure 7.9 shows both the FMM 3-segment solution and the ideal knee of MMSEA solutions. The ideal knee is the solution that has the shortest distance to the (usually impossible) ideal solution of the problem. The ideal solution is defined as a solution that has the lowest value of both objectives.

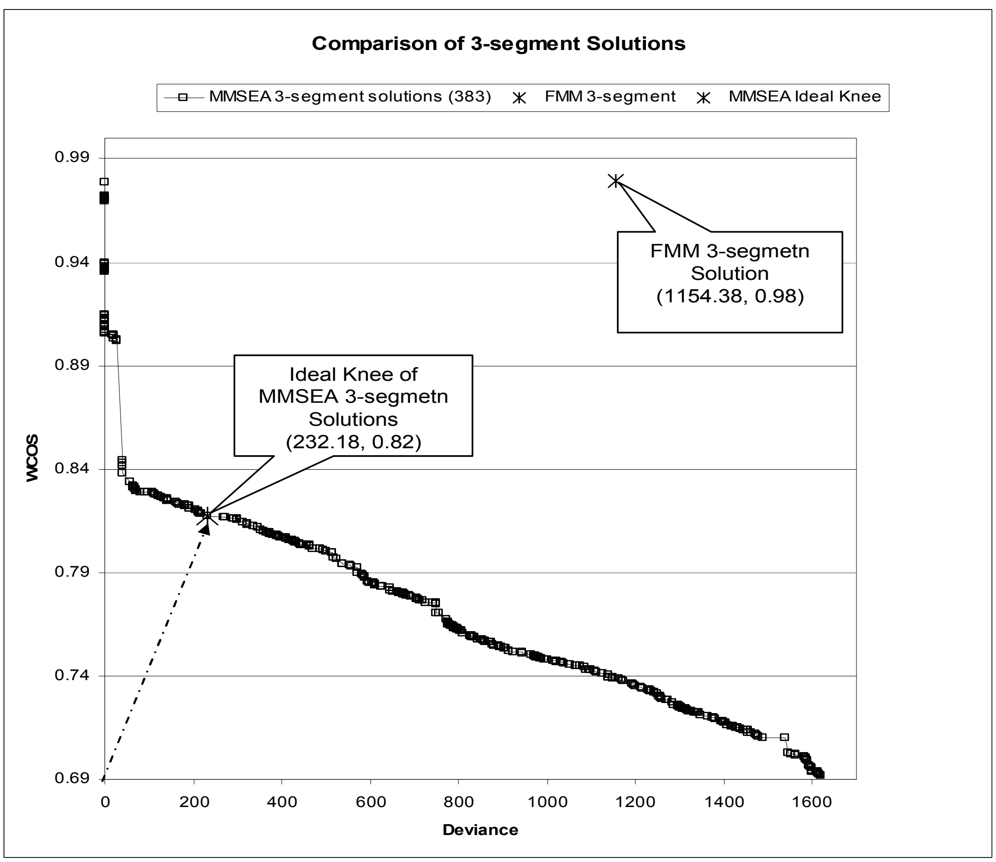


Figure 7.9 Comparison of 3-segment solutions

	MMSEA Segment		FMM Segment	
Segment Size	593		1346	
Variable	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
Age	44.33	12.33	46.45	12.65
Children	0.65	0.48	0.58	0.49
Income	6.43	2.07	6.36	2.03
DSLP	76.28	53.21	44.29	44.98
Margin	17.63	19.24	14.95	14.49
Purchase	2.89	2.55	5.15	4.49
Target	0.06	0.24	0.60	0.49

Table 7.7 Comparison of segment 1

	MMSEA Segment		FMM Segment	
Segment Size	491		95	
Variable	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
Age	44.31	10.51	48.21	11.87
Children	.99	.11	.64	.48
Income	6.71	1.87	6.49	2.12
DSLP	29.44	29.47	101.66	43.99
Margin	14.00	11.86	17.13	19.62
Purchase	6.12	4.82	2.28	0.96
Target	0.98	0.13	0.93	0.26

Table 7.8 Comparison of segment 2

	MMSEA Segment		FMM Segment	
Segment Size	416		59	
Variable	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
Age	51.79	13.66	42.10	12.16
Children	0.02	0.13	0.64	0.48
Income	5.81	2.15	5.90	2.51
DSLP	31.11	32.89	57.12	44.13
Margin	13.05	9.38	16.93	14.00
Purchase	6.46	4.84	4.39	4.38
Target	0.99	0.11	0.49	0.50

Table 7.9 Comparison of segment 3

Table 7.7, Table 7.8, and Table 7.9 are the profiles of the corresponding segment 1, 2, and 3 of the 3-segment solutions of MMSEA and FMM. Segments are compared based on their size, i.e., segment 1 is the biggest while segment 3 is the smallest. There are some interesting observations from these segment profiles. First, the MMSEA segments have similar sizes. The two small segments of the FMM solution only have about 10% of the customers. This could be a big concern because the size might be too small for a marketing campaign. Second, MMSEA segments are more homogeneous than FMM segments because the variable standard deviations of MMSEA are usually smaller than that of FMM. The predictor means of MMSEA segments are significantly different from each other (Table 7.10) while FMM segments are not significantly different from each other (Table 7.11).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	16755.715	2	8377.858	56.596	.000
	Within Groups	221600.218	1497	148.030		
	Total	238355.933	1499			
Children	Between Groups	216.526	2	108.263	1098.401	.000
	Within Groups	147.551	1497	.099		
	Total	364.077	1499			
Income	Between Groups	191.349	2	95.675	23.229	.000
	Within Groups	6165.901	1497	4.119		
	Total	6357.250	1499			
DSLP	Between Groups	761883.376	2	380941.688	223.578	.000
	Within Groups	2550648.960	1497	1703.840		
	Total	3312532.336	1499			
Margin	Between Groups	6128.751	2	3064.375	14.139	.000
	Within Groups	324451.182	1497	216.734		
	Total	330579.933	1499			
Purchase	Between Groups	4143.477	2	2071.738	124.254	.000
	Within Groups	24960.002	1497	16.673		
	Total	29103.479	1499			

Table 7.10 One-way ANOVA of MMSEA Solution

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	1405.203	2	702.601	4.439	.012
	Within Groups	236950.731	1497	158.284		
	Total	238355.933	1499			
Children	Between Groups	.568	2	.284	1.170	.311
	Within Groups	363.509	1497	.243		
	Total	364.077	1499			
Income	Between Groups	14.151	2	7.076	1.670	.189
	Within Groups	6343.099	1497	4.237		
	Total	6357.250	1499			
DSLP	Between Groups	296784.497	2	148392.248	73.661	.000
	Within Groups	3015747.839	1497	2014.528		
	Total	3312532.336	1499			
Margin	Between Groups	611.670	2	305.835	1.388	.250
	Within Groups	329968.263	1497	220.420		
	Total	330579.933	1499			
Purchase	Between Groups	747.735	2	373.868	19.738	.000
	Within Groups	28355.744	1497	18.942		
	Total	29103.479	1499			

Table 7.11One-way ANOVA of FMM Solution

Third, it's easy to formulate a marketing campaign base on the MMSEA solution because of the improved identifiability and responsiveness. Both segment 2 and segment 3 of the MMSEA solution have high response rates (0.98 and 0.99). They are significantly different from the low responsiveness in segment 1. Segment 2 is characterized by younger age, higher children presence rate, above average income, large purchase number and higher margin. Segment 3 is characterized by elder age, lower children presence rate, lower income level, large purchase number and lower margin.

7.4.3 Predictive Power

One interesting question for the predictive logistic model is what the prediction performance of the solution is. We developed an R program to calculate the prediction correctness of 10,743 non-sample customers based on the segment properties of 1,500 sample customers. The basic program logic is shown in Figure 7.10.

```
Initialize subCount, totalCount to 0

For each segment of sample customers
        Create the logistic regression model
        Calculate segment center

For each customer in non-sample customers
        Calculate the distance to each sample segment center
        Assign its segment number to the number of the nearest segment
        center

For each segment in non-sample customers
        Use the corresponding sample logistic model to predict its response
        flag
        Set the subCount to the number of correctly predicted customers
        Set the segment prediction correctness rate as the ratio of subCount
        and segment size
        Add subCount to totalCount

Set the overall prediction correctness rate as the ratio of totolCount and
non-sample customer size
```

Figure 7.10 Code for calculating prediction correction rate

Because the selected MMSEA solution has better predictive power and within segment homogeneity, we expect that if we segment other non-sample customers according to the segment profile of sample customers, the predictive power of the MMSEA solution should be higher than that of FMM. The results of selected 3-segment MMSEA and FMM solutions are shown in Table 7.12. The MMSEA segments are more homogeneous and two out of three have better prediction performance. The overall prediction performance of the MMSEA solution is also better than the overall prediction performance of the FMM solution.

	MMSEA		**FMM**	
Deviance	231.19		1154.38	
WCOS	0.82		0.98	
	Sample (Size)	Test(Size)	Sample (size)	Test (size)
Overall prediction correctness				
	0.98 (1463/1500)	0.68(7325/10743)	0.82 (1228/1500)	0.57 (6184/10743)
Segment level correctness				
1	0.94 (558/593)	0.64(1847/2894)	0.80(1074/1346)	0.74(4283/5791)
2	1.00 (489/491)	0.68(2899/4241)	1.00 (95/95)	0.34(791/2296)
3	1.00 (416/416)	0.71(2579/3608)	1.00 (59/59)	0.40(1074/2656)

Table 7.12 Comparison of Prediction Accuracy

We also investigated the prediction performance of all MMSEA solutions. The question is whether there is a relationship between the prediction performance and solution objectives (WCOS and Deviance). Figure 7.11 to Figure 7.16 show the results for the 3-segment to 5-segment solutions. The prediction performance is rather independent of the WCOS and Deviance objectives except in a few extreme cases. However, the prediction performance can vary significantly even if the WCOS or Deviance objectives are similar. It may be worth finding those solutions with high prediction performance within the small range of given optimization objectives.

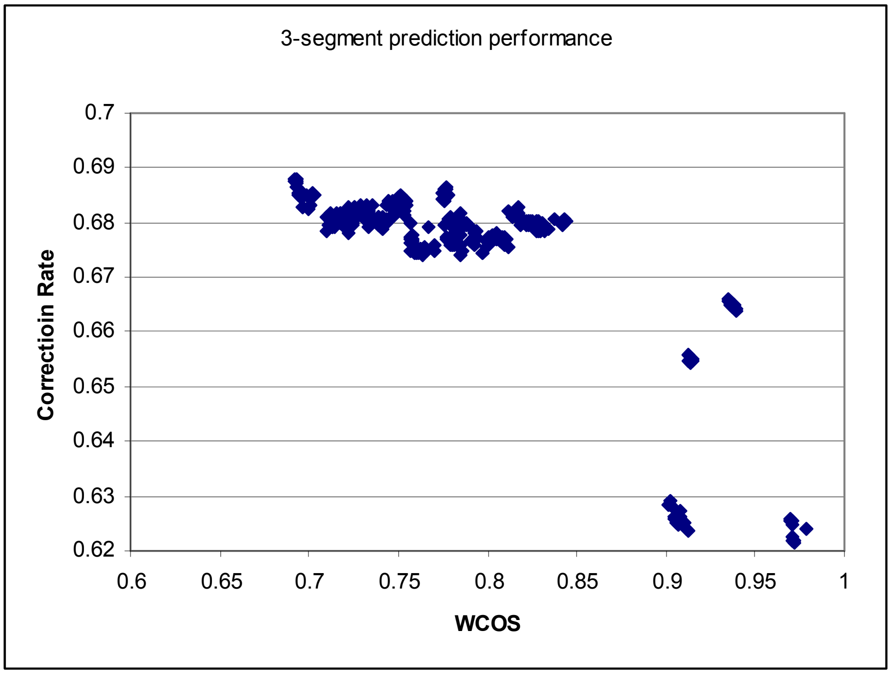


Figure 7.11 3-segment prediction performance vs. WCOS

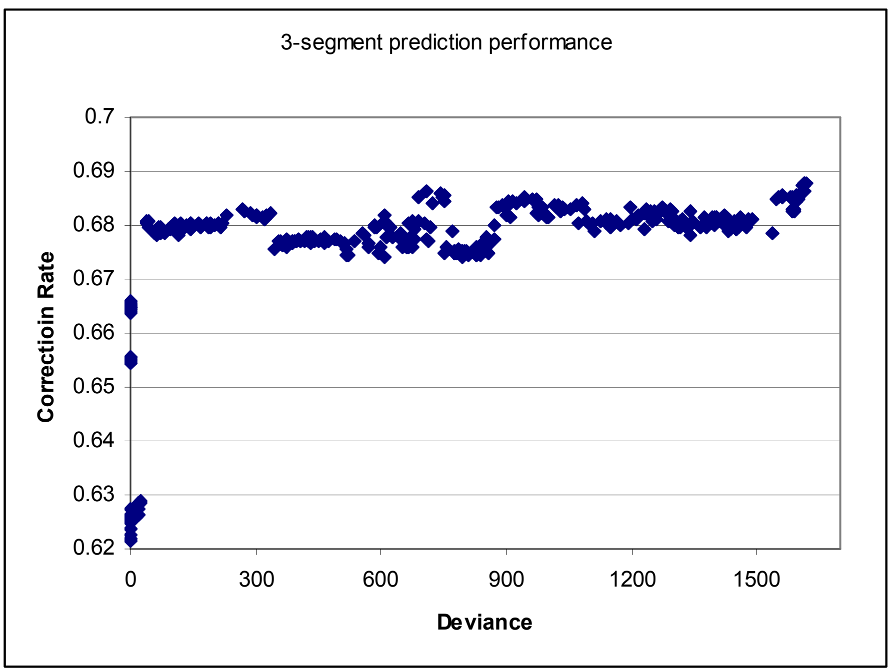


Figure 7.12 3-segment prediction performance vs. deviance

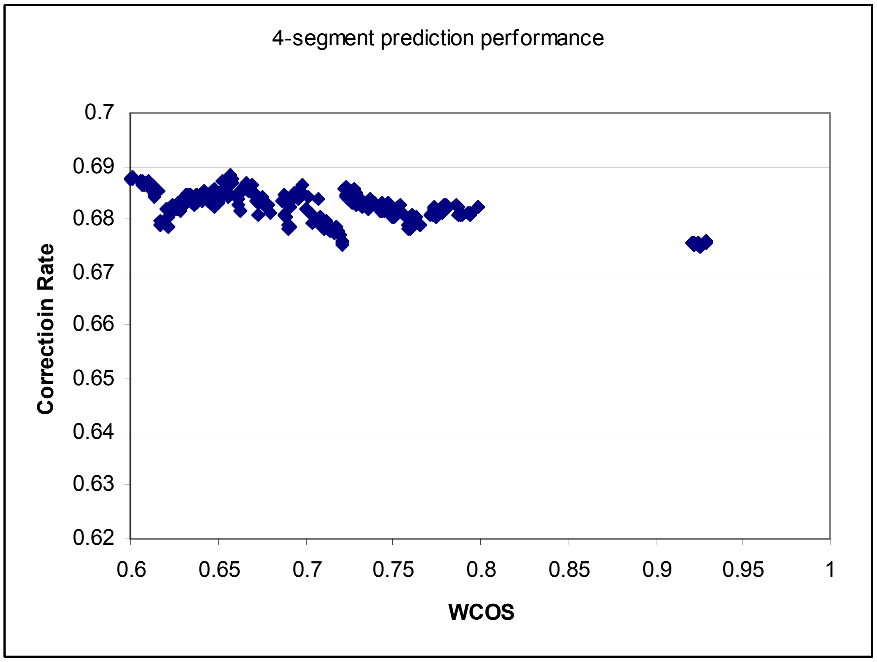


Figure 7.13 4-segment prediction performance vs. WCOS

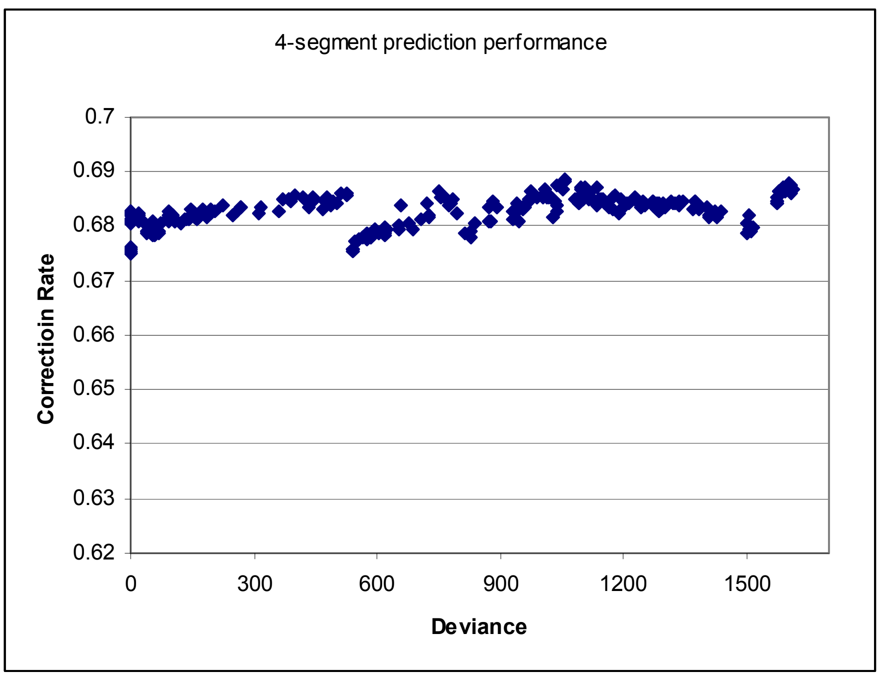


Figure 7.14 4-segment prediction performance vs. deviance

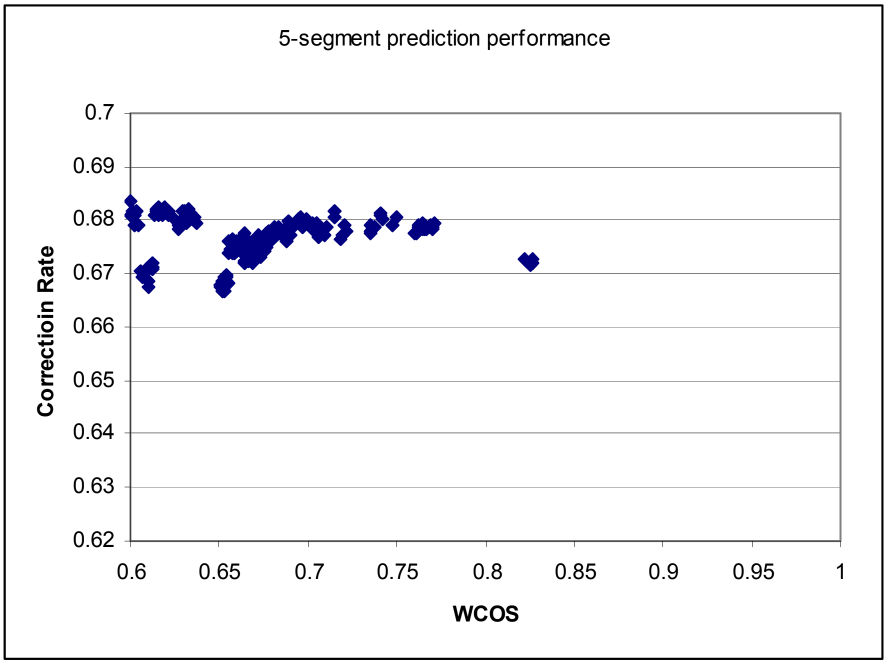


Figure 7.15 5-segment prediction performance vs. WCOS

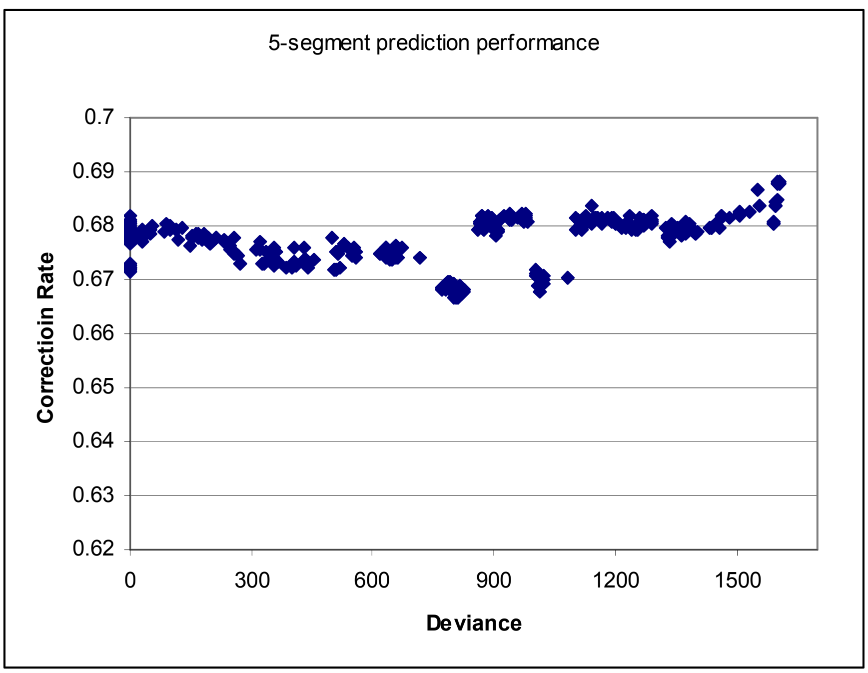


Figure 7.16 5-segment prediction performance vs. deviance

Because firms often target selected segment(s), it's also interesting to investigate the relationships between the best predicted segment and segment optimization objectives. As the overall prediction performance, there is no obvious relationship between best segment prediction performance and optimization objectives except in some extreme cases. However, it is expected that the best segment prediction performance increases as the number-of-segments increases because a small segment usually has better predictive power and higher within segment homogeneity. The results are shown in Figure 7.17 to Figure 7.22. The 3-segment best prediction correction rate is 0.707, 4-segment is 0.724 and 5-segment is the best with a value of 0.788. Each value is the average of all best segment prediction correction rates.

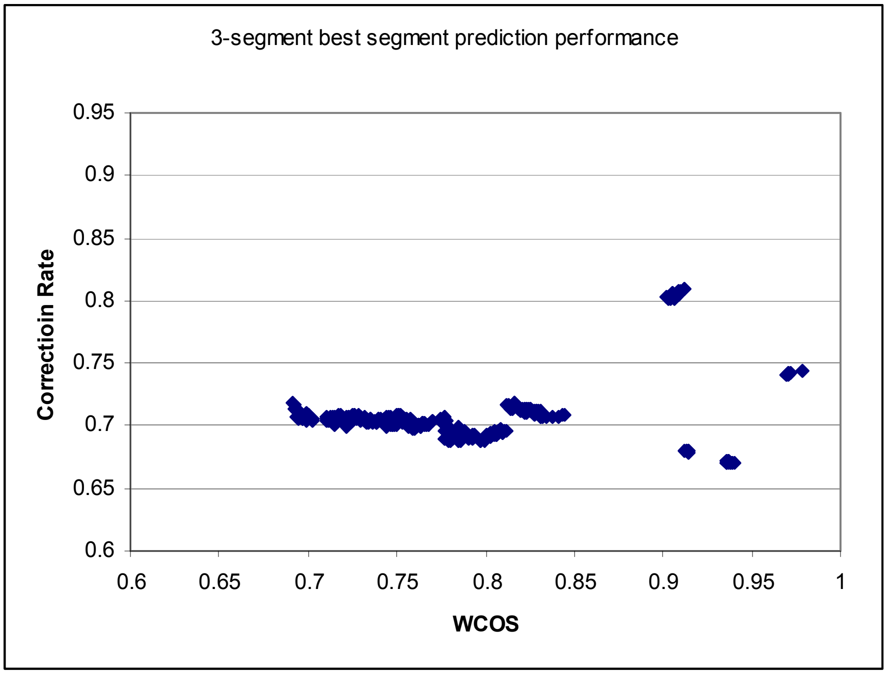


Figure 7.17 3-segment best segment prediction performance vs. WCOS

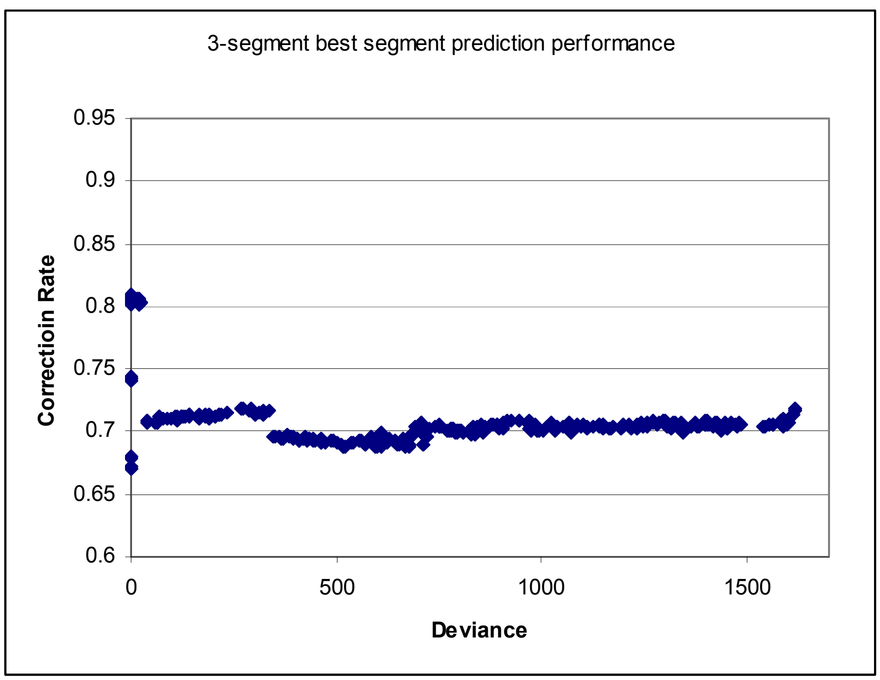


Figure 7.18 3-segment best segment prediction performance vs. Deviance

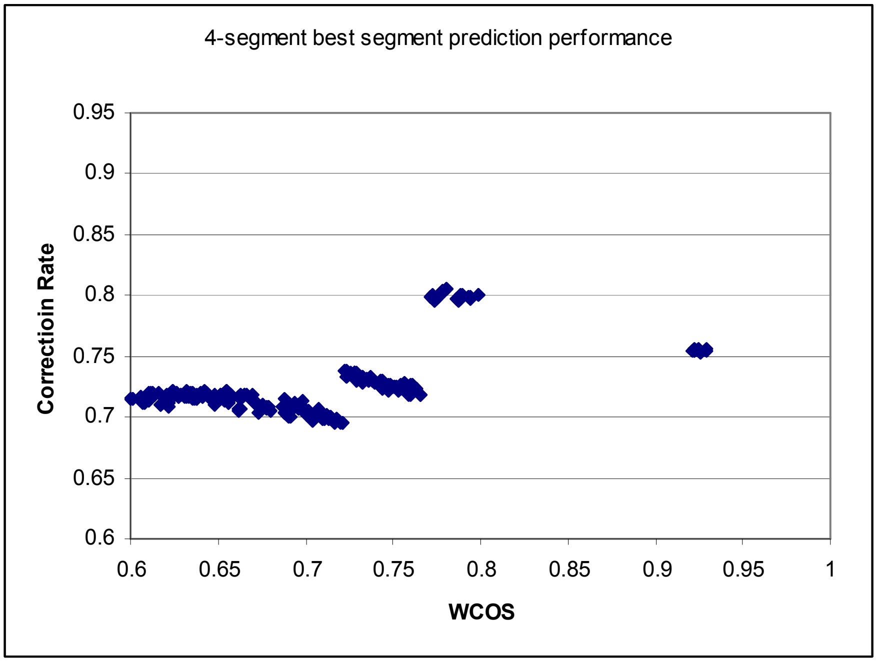


Figure 7.19 4-segment best segment prediction performance vs. WCOS

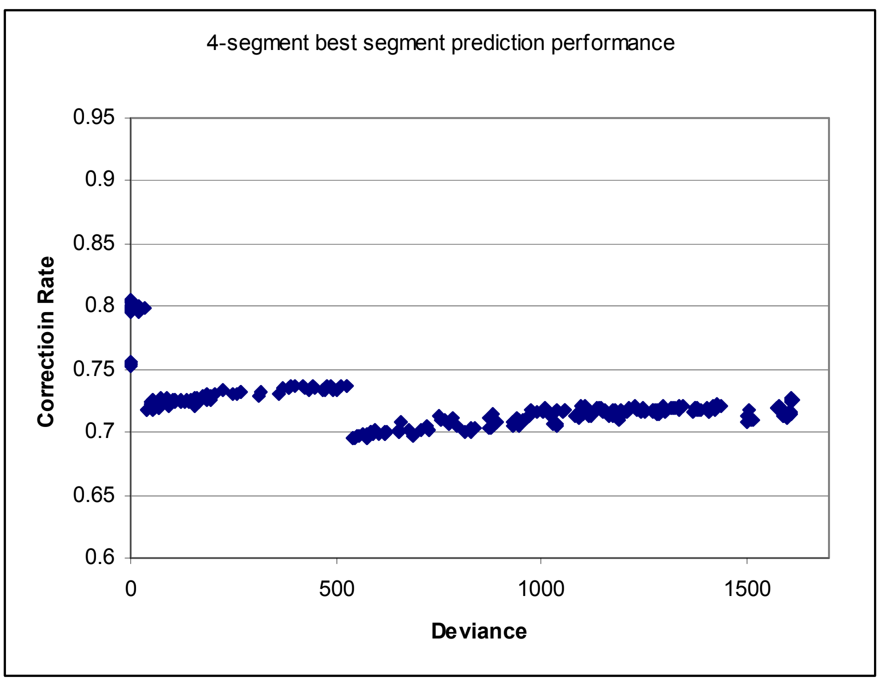


Figure 7.20 4-segment best segment prediction performance vs. Deviance

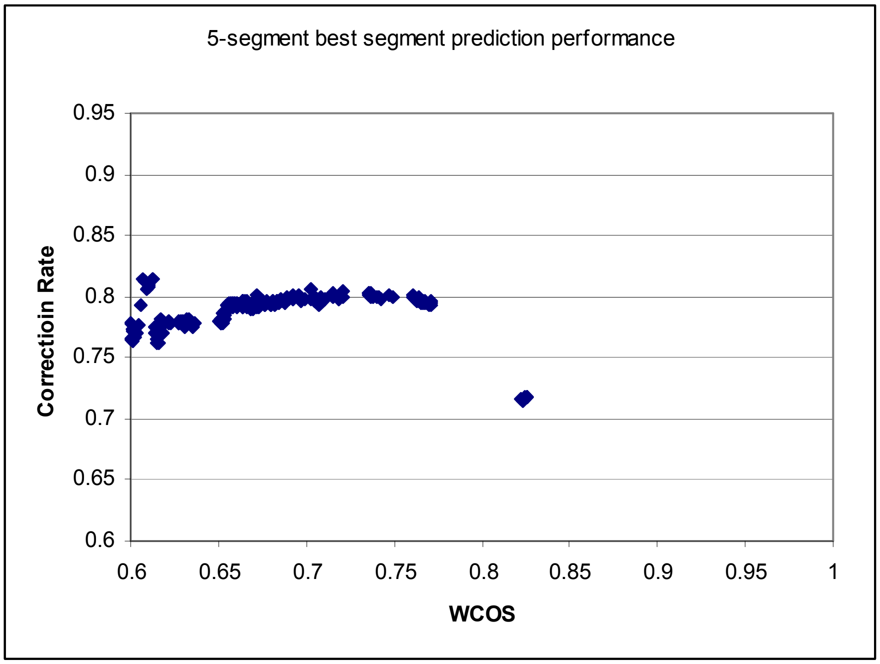


Figure 7.21 5-segment best segment prediction performance vs. WCOS

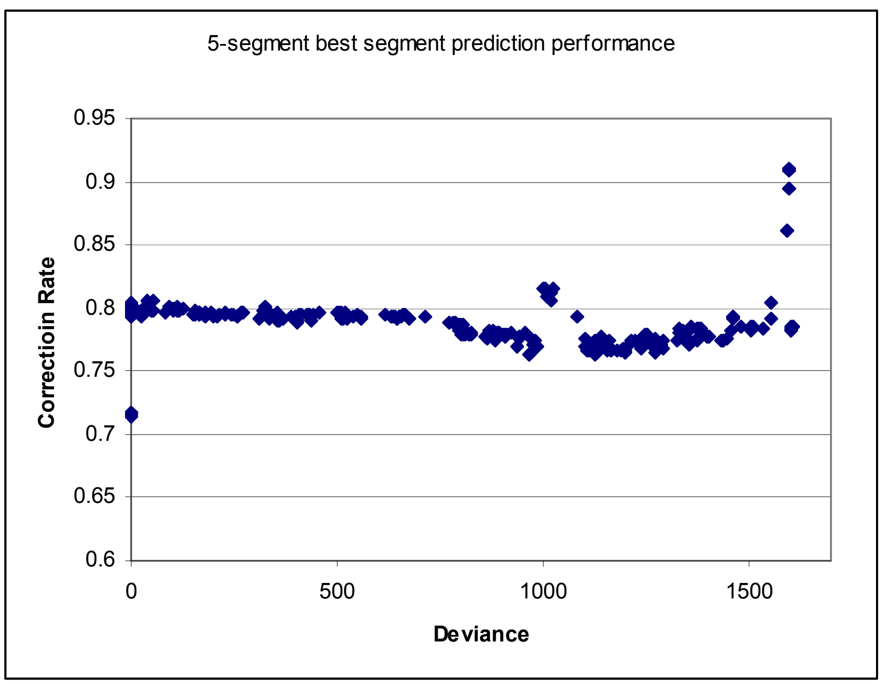


Figure 7.22 5-segment best segment prediction performance vs. Deviance

7.5 Evaluation Conclusion

The evaluation of the logistic regression finds similar results to the previous chapter. Compared with the FMM method, the MMSEA algorithm not only can give a holistic view of the solutions, but it also can find better solutions that are optimized in both objectives. The MMSEA is able to generate the Pareto front from a few solutions initialized from clusterwise logistic regression and K-means methods.

The prediction performance is another dimension that decision makers should investigate. This research finds that the prediction performance is independent of the optimization objectives of within segment heterogeneity and predictive power. However, the best segment prediction performance depends on the number-of-segments.

The logistic regression results raise some issues to solve. The MMSEA algorithm in its current version is severely limited by computer memory. In a 32 bit operating system like Windows XP, it can only process several thousand customers. Consequently, there are many segments having "perfect" predictive power. We attribute this "perfection" to the small sample size and it hurts the prediction performance on non-sample data because logistic regression does not converge properly for such data. We expect that the prediction performance increases when the sample size increases. This is an issue to be solved in future research.

The performance of MMSEA is summarized in Table 7.13. The MMSEA performance can be improved significantly with a parallel algorithm running in multiple CPUs. The evolutionary algorithm can be easily transformed into a parallel algorithm.

FMM (5 runs)	Each MMSEA Generation	Clusterwise Regression	K-means
455.7 seconds	17.5 seconds (40 candidates)	1.5 hour	< 1 second

Table 7.13 Runtime performance

In the above analysis, the 991 Pareto optimal solutions are generated in about 5 days, 22 hours.

8 CONCLUSION

After discussing the multicriterion nature of market segmentation, this dissertation proposed a unified market segmentation model based on multiobjective optimization algorithms. To the best of our knowledge, this is the first of a kind that defines the market segmentation problem using the multiobjective optimization notation based on Pareto optimality and a set of acceptable solutions.

8.1 Research Evaluation According to the Design Science Guidelines

The research methodology of this dissertation follows the design science research guidelines. Consequently, this research can be evaluated according to the seven guidelines proposed by Hevner et al (Hevner, March et al. 2004). The evaluation is shown in Table 8.1.

Guideline	Evaluation
1. Design as an Artifact	This research produces a new conceptual model and solution techniques as testable artifacts.
2. Problem Relevance	This research develops new mathematic model and algorithms to the very important and relevant market segmentation problem.
3. Design Evaluation	The research is both theoretically analyzed and empirically evaluated using three real business data sets.
4. Research Contribution	The research brings new perspectives, methodology and solution techniques to the market segmentation problem.
5. Research Rigor	The market segmentation is abstracted and formulated as a multiobjective optimization problem. The results are empirically tested using statistical methods.
6. Design as a Search Process	The research is a result of three years iterative design process. The proposed model and solution techniques integrate knowledge from both the business domain and information technology.
7. Communication of Research	The research is presented with rigorous theoretical analysis and empirical evaluation in real business setting. The important algorithm design decisions are clearly discussed.

Table 8.1 Research evaluation according to the design science guidelines

8.2 Discussion and Future Research Directions

The proposed model considers multiple objectives (criteria) simultaneously in searching for global optimal solutions. Consequently, the model generalizes existing single objective methods as well as unifies the descriptive and predictive market segmentation methods. Another distinct advantage resulting from this multiobjective view is that a set of diverse Pareto optimal solutions for a range of number-of-segments is generated. As a result, no multicriterion aggregation or tradeoffs are required before the users see the results. It's well known that those a priori decisions are very difficult to make and usually put unpredictable constraints on searching for globally optimal solution(s). The set of solutions over a range of number-of-cluster values provides a holistic view that gives decision makers a lot of flexibility and insights that are missing in most existing methods. Additionally, Pareto optimality avoids the non-commensurable issues in many market segmentation problems.

As a proven meta-heuristic technique, multiobjective evolutionary algorithms search for global optimal solutions and are robust in terms of data types and objective function forms (discrete or continuous, concave or convex, single modal or multimodal). These features make them a good implementation for the multiobjective optimization component of the unified market segmentation model. As shown by the constraint of minimum segment size, the MMSEA can incorporate constraints into the search process. Different manual, automatic or mixed methods could be used for the solution analysis part of the model.

We have evaluated the unified market segment and the MMSEA implementation in three cases – all are with real data. In joint market segmentation, MMSEA can give a holistic view of the possible solutions. In multicriterion ordinary linear regression and logistic regression segmentation, the MMSEA method is compared with the FMM method. Not only can MMSEA find better solutions for both descriptive and responsiveness criteria, but MMSEA solutions also have more evenly distributed segments. The segment profile and optimization objectives in both predictive power and segment homogeneity give decision makers much desired flexibility in selecting the most appropriate solution for a specific market program.

Nonetheless, this unified market segmentation model and the multiobjective evolutionary-based implementation have many research issues to be solved. First, the performance of MMSEA could be improved by using parallel program techniques and different gene representations. Second, as a meta-heuristic method, MMSEA could not give the confidence level of the optimization result while the statistic model-based approaches usually provide this useful information. How to extend the multiobjective model to provide the desired statistic property is a question to be answered to fully take advantage of the benefit of this unified market segmentation method. Third, the multiobjective evolutionary algorithm is just one of many kinds of multiobjective optimization methods such as tabu search, simulated annealing, and memetic algorithms. The appropriateness, efficiency, and effectiveness of other multiobjective optimization methods in multiobjective market segmentation are to be studied. Fourth, how the solution technique integrates the preferences and constraints before, during and after the multiobjective optimization and how the solution technique effectively interact with market segmentation decision makers are interesting research topic. Finally but equally important, the multiobjective analytic part doesn't have a sound theoretical foundation to support the objective selection of best solution from a set of Pareto optimal solutions. If this issue is solved, many interesting research topics such as the evolution of market structure could be initiated with a solid base.

APPENDICES

APPENDIX A DATA PREPARATION ONE

A.1 Introduction

There are two types of data: 1) the purchase data in year 2005, from approximately Jan to Dec. The number of records is 50,000. 2) The household level demographic data of each customer. The number of records is 43,340. The data is linked by a master customer identifier. The data description is shown in Table A.1.

Field Name	**Field Description**
CustID	Customer master identifier.
Target	Flag of whether the customer responds to the marketing campaign.
DSLP	Days since last purchase.
ADBP	Average days between purchases.
AVG_NET_SALES_12MO	Average net sales over the past 12 months.
AVG_NET_SALES_6MO	Average net sales over the past 6 months.
AVG_NET_SALES_3MO	Average net sales over the past 3 months.
NUM_PURCH_12_MO	Number of purchase over the past 12 months.
NUM_PURCH_6_MO	Number of purchase over the past 6 months.
NUM_PURCH_3_MO	Number of purchase over the past 3 months.
AVG_MGN_12MO	Average margin over the past 12 months.
AVG_MGN_6MO	Average margin over the past 6 months.
AVG_MGN_3MO	Average margin over the past 3 months.
ITEM_12MO	Average number of items per transaction over the past 12 months.
ITEM_3MO	Average number of items per transaction over the past 3 months.
TIRE_NEW	Flag of whether the customer is new in the past 6 months.
TIRE_PLATINUM	Customer Segment: Platinum. Customer has 4 or more transactions in past 6 months with an average purchase of >= $35.
TIRE_GOLD	Customer Segment: Gold. Customer has 4 or more transactions in past 6 months with an average purchase of < $35.
TIRE_SILVER	Customer Segment: Silver. Customer has 1 to 3 transactions in past 6 months with an average purchase >= $35.
TIRE_BRONZE	Customer Segment: Bronze. Customer has 1 to 3 transactions in past 6 months with an average purchase of < $35.

Field Name	Field Description
Age	AGE IN TWO-YEAR INCREMENTS - INPUT INDIVIDUAL – PREMIER 17 = AGE LESS THAN 18 18 = AGE 18 - 19 20 = AGE 20 - 21 22 = AGE 22 - 23 24 = AGE 24 - 25 26 = AGE 26 - 27 28 = AGE 28 - 29 30 = AGE 30 - 31 32 = AGE 32 - 33 34 = AGE 34 - 35 36 = AGE 36 - 37 38 = AGE 38 - 39 40 = AGE 40 - 41 42 = AGE 42 - 43 44 = AGE 44 - 45 46 = AGE 46 - 47 48 = AGE 48 - 49 50 = AGE 50 - 51 52 = AGE 52 - 53 54 = AGE 54 - 55 56 = AGE 56 - 57 58 = AGE 58 - 59 60 = AGE 60 - 61 62 = AGE 62 - 63 64 = AGE 64 - 65 66 = AGE 66 - 67 68 = AGE 68 - 69 70 = AGE 70 - 71 72 = AGE 72 - 73 74 = AGE 74 - 75 76 = AGE 76 - 77 78 = AGE 78 - 79 80 = AGE 80 - 81 82 = AGE 82 - 83 84 = AGE 84 - 85 86 = AGE 86 - 87 88 = AGE 88 - 89 90 = AGE 90 - 91 92 = AGE 92 - 93 94 = AGE 94 - 95 96 = AGE 96 - 97 98 = AGE 98 - 99 99 = AGE GREATER THAN 99 DEFAULT IS BLANK(S)
Gender	GENDER - INPUT INDIVIDUAL – PREMIER M = MALE F = FEMALE DEFAULT IS BLANK(S)

Field Name	Field Description
Occupation	OCCUPATION - INPUT INDIVIDUAL – PREMIER 1 = PROFESSIONAL / TECHNICAL 2 = ADMINISTRATION / MANAGERIAL 3 = SALES / SERVICE 4 = CLERICAL / WHITE COLLAR 5 = CRAFTSMAN / BLUE COLLAR 6 = STUDENT 7 = HOMEMAKER 8 = RETIRED 9 = FARMER A = MILITARY B = RELIGIOUS C = SELF EMPLOYED D = SELF EMPLOYED - PROFESSIONAL / TECHNICAL E = SELF EMPLOYED - ADMINISTRATION / MANAGERIAL F = SELF EMPLOYED - SALES / SERVICE G = SELF EMPLOYED - CLERICAL / WHITE COLLAR H = SELF EMPLOYED - CRAFTSMAN / BLUE COLLAR I = SELF EMPLOYED - STUDENT J = SELF EMPLOYED - HOMEMAKER K = SELF EMPLOYED - RETIRED L = SELF EMPLOYED - OTHER V = EDUCATOR W = FINANCIAL PROFESSIONAL X = LEGAL PROFESSIONAL Y = MEDICAL PROFESSIONAL Z = OTHER DEFAULT IS BLANK(S)
Education	EDUCATION INPUT INDIVIDUAL – PREMIER 1 = COMPLETED HIGH SCHOOL 2 = COMPLETED COLLEGE 3 = COMPLETED GRADUATE SCHOOL 4 = ATTENDED VOCATIONAL/TECHNICAL DEFAULT IS BLANK(S)
Marital	MARITAL STATUS IN THE HOUSEHOLD – PREMIER M = MARRIED S = SINGLE A = INFERRED MARRIED B = INFERRED SINGLE DEFAULT IS BLANK(S)
WorkingWoman	WORKING WOMAN – PREMIER Y = WORKING WOMAN - PREMIER DEFAULT IS BLANK(S)
Children	CHILDREN - PRESENCE IN HOUSEHOLD – PREMIER Y = CHILD(REN) PRESENT N = NO CHILDREN PRESENT DEFAULT IS BLANK(S)

Field Name	Field Description
Adults	ADULTS - NUMBER IN HOUSEHOLD – PREMIER 1 = 1 ADULT 2 = 2 ADULTS 3 = 3 ADULTS 4 = 4 ADULTS 5 = 5 ADULTS 6 = GREATER THAN 5 ADULTS DEFAULT IS BLANK(S)
CreditCard	CREDIT CARD INDICATOR – PREMIER BANK CARD HOLDER 1 = TRUE DEFAULT IS 0 GAS/DEPARTMENT/RETAIL CARD HOLDER 1 = TRUE DEFAULT IS 0 TRAVEL AND ENTERTAINMENT CARD HOLDER 1 = TRUE DEFAULT IS 0 CREDIT CARD HOLDER - UNKNOWN TYPE 1 = TRUE DEFAULT IS 0 PREMIUM CARD HOLDER 1 = TRUE DEFAULT IS 0 UPSCALE (DEPARTMENT STORE) CARD HOLDER 1 = TRUE DEFAULT IS 0
Income	INCOME CODE - ESTIMATED HOUSEHOLD – PREMIER 1 = LESS THAN $15,000 2 = $15,000 - $19,999 3 = $20,000 - $29,999 4 = $30,000 - $39,999 5 = $40,000 - $49,999 6 = $50,000 - $74,999 7 = $75,000 - $99,999 8 = $100,000 - $124,999 9 = GREATER THAN $124,999 DEFAULT IS BLANK(S)
Homeowner	HOME OWNER / RENTER – PREMIER O = HOME OWNER R = RENTER DEFAULT IS BLANK(S)

Appendix Table A.1 Description of All Variables

We merged the two files using master customer identifier and created a data file with 43,340 customers. Because there are 22,419 new customers are new in past 6 months, we removed theses new customers and used the purchase data of the past 6 months. In other words, we only analyzed customers that had been in the premium club for at least 6 months. There are 20,921 such customers.

A.2 Data Cleaning and Coding

We removed the following variables from the analysis:

- Occupation: there are 25 categories -- too many for the algorithm.
- Credit card: there are 6 types and one customer may have more than one type. It is difficult to code and interpret these data.
- Purchase/Sales/Margin 12MO and 3 MO: as explained above, only 6 month data is used.
- Tier new/lapsed/platinum/gold/silver/bronze: these are constants or derived data.

There is no average item per transaction data for the 6 month period. Comparing the 12 month data and the 3 month data, we chose the 12month data as the average number of items per transaction because it is more stable than the 3 month data. For example, 3,820 customers (about 19%) have 0 items/transactions in the past 3 months.

We recoded using the following rules:

Gender: female = 1, male = 0

Marital: M or A =1, S | B = 0

WorkingWoman: Y = 1, else = 0 (there is N sign, but given that 34% marked as Y, we would guess the rest are "N").

Children: Y =1, N = 0

Homeowner: O =1, R = 0

The resulting attributes are numeric. The statistics are as follows.

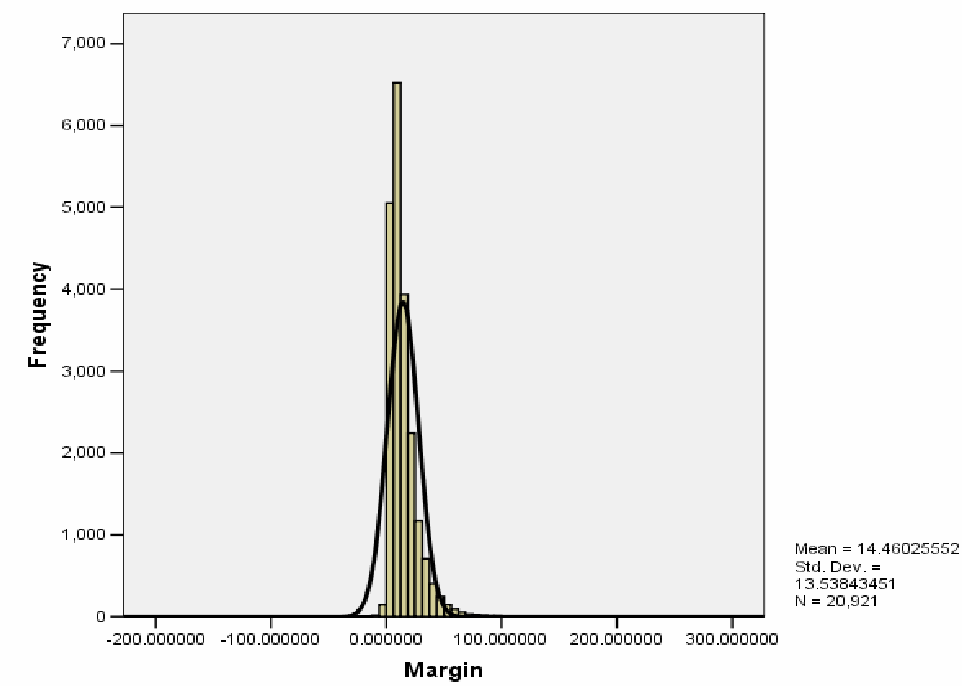
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Target	20921	.00	1.00	.60	.49
Age	18843	18.00	99.00	45.67	13.59
Gender	19526	.00	1.00	.72	.45
Education	8362	1.00	4.00	1.61	.78
Marital	18587	.00	1.00	.68	.47
WorkingWoman	20921	.00	1.00	.34	.47
Children	12872	.00	1.00	.59	.49
Adult	20048	1.00	6.00	1.96	.77
Income	19421	1.00	9.00	6.08	2.15
HomeOwner	16664	.00	1.00	.97	.17
DSLP	20921	.00	299.00	47.86	47.45
ADBP	20790	-1.00	180.50	34.57	22.22
Sales	20921	-169.98	379.99	31.26	22.66
Margin	20921	-107.59	283.08	14.46	13.54
Purchase	20921	.00	177.00	5.20	5.37
Item	20921	.50	436.17	7.19	12.02
Valid N (listwise)	5966				

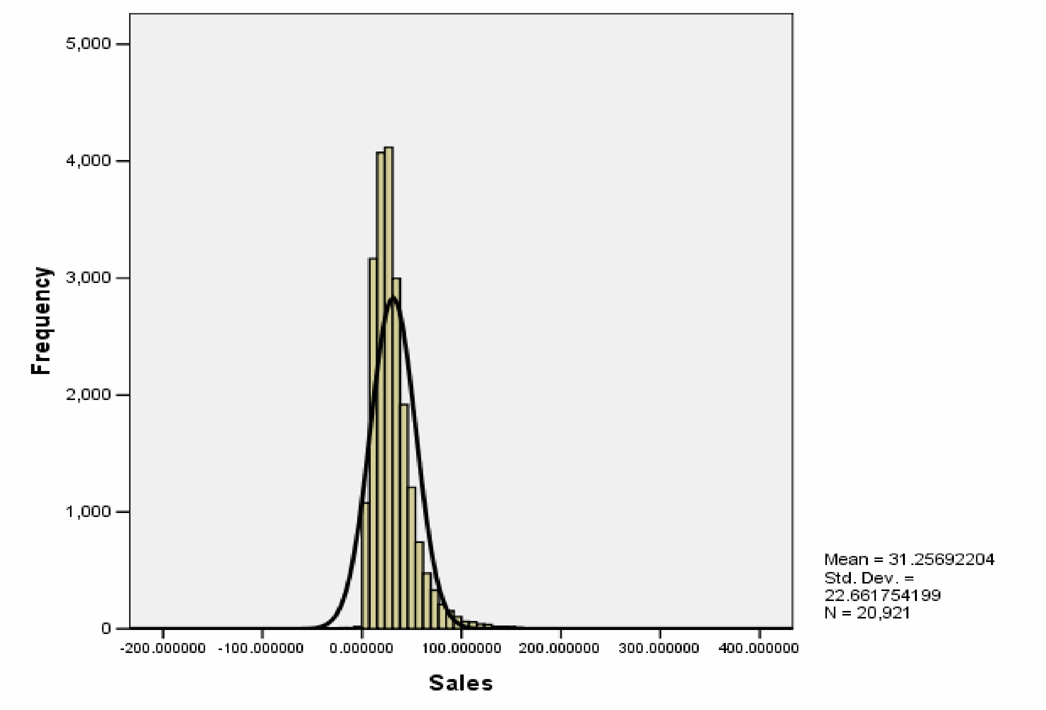
Appendix Table A.2 Descriptive Statistics of Cleaned Data

There are some interesting things in the above statistics: 1) 97% of the customers are home owner. 2) Education level has a high missing value rate (60%). So we took the home owner and education out. Additionally, response data (target) is not used in this model.

The margin is not normally distributed. There are 455 cases whose margin values are zero or negative. Those values are skewed and may represent abnormal issues. They could be a special segment to be studied.



Appendix Figure A.1 Histogram of Margin



Appendix Figure A.2 Histogram of Sales

It turned out that all purchase data (the SDLP, ADBP, Sale, Margin, Purchase and Item) are highly skewed.

Descriptive Statistics

	N		Skewness	
	Statistic	Std. Error	Statistic	Std. Error
Age	18843		.347	.018
Gender	19526		-.964	.018
Marital	18587		-.762	.018
WorkingWoman	20921		.674	.017
Children	12872		-.372	.022
Adult	20048		.897	.017
Income	19421		-.592	.018
DSLP	20921		1.227	.017
ADBP	20790		1.781	.017
Sales	20921		2.939	.017
Margin	20921		3.467	.017
Purchase	20921		6.791	.017
Item	20921		8.638	.017
Valid N (listwise)	11871			

Appendix Table A.3 Descriptive statistics of all remaining variables

So we used ln() transformation on all purchase data. As a result, zero and negative values are excluded from the result during the transformation because the clustering algorithm needs standard values. All data are standardized. We run model selection before filtering out cases with missing values because removing variables first results in more cases without missing values.

A.3 Model Selection

The model is to predict customer sales or profitability from other variables. The stepwise regression results are similar but margin is more interesting to marketers. So we chose margins as the dependent variable.

Model Summary

Model	R	R Square[a]	Adjusted R Square	Std. Error of the Estimate
1	.149[b]	.022	.022	.98292458
2	.191[c]	.036	.036	.97589492
3	.216[d]	.047	.046	.97073624
4	.218[e]	.048	.047	.97029147
5	.220[f]	.048	.048	.96988251
6	.221[g]	.049	.048	.96974973
7	.225[h]	.051	.050	.96882838
8	.226[i]	.051	.051	.96859592

a. For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This CANNOT be compared to R Square for models which include an intercept.

b. Predictors: Zscore(LnItem)

c. Predictors: Zscore(LnItem), Zscore(Income)

d. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age)

e. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender)

f. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender), Zscore(Marital)

g. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender), Zscore(Marital), Zscore(LnPurchase)

h. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender), Zscore(Marital), Zscore(LnPurchase), Zscore(LnADBP)

i. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender), Zscore(Marital), Zscore(LnPurchase), Zscore(LnADBP), Zscore(LnDSLP)

Appendix Table A.4 Stepwise regression of Margin

Model Summary

Model	R	R Square[a]	Adjusted R Square	Std. Error of the Estimate
1	.269[b]	.072	.072	.95484292
2	.307[c]	.094	.094	.94349385
3	.311[d]	.097	.097	.94227433
4	.313[e]	.098	.098	.94163321
5	.316[f]	.100	.099	.94096861
6	.319[g]	.102	.102	.93971382
7	.320[h]	.103	.102	.93948999
8	.321[i]	.103	.102	.93933490

a. For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This CANNOT be compared to R Square for models which include an intercept.

b. Predictors: Zscore(LnItem)

c. Predictors: Zscore(LnItem), Zscore(Income)

d. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age)

e. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender)

f. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender), Zscore(LnADBP)

g. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender), Zscore(LnADBP), Zscore(LnPurchase)

h. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender), Zscore(LnADBP), Zscore(LnPurchase), Zscore(Marital)

i. Predictors: Zscore(LnItem), Zscore(Income), Zscore(Age), Zscore(Gender), Zscore(LnADBP), Zscore(LnPurchase), Zscore(Marital), Zscore(WorkingWoman)

Appendix Table A.5 Stepwise regression of Sales

The interesting thing is that 1) the difference between margin and sales is because of the explanation power of items. 2) Purchase data only explains about 0.02% R squares of margin.

Model Summary

Model	R	R Square[a]	Adjusted R Square	Std. Error of the Estimate
1	.113[b]	.013	.013	.98878041
2	.142[c]	.020	.020	.98513964
3	.147[d]	.022	.021	.98439996
4	.151[e]	.023	.023	.98379373

a. For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This CANNOT be compared to R Square for models which include an intercept.

b. Predictors: Zscore(Income)

c. Predictors: Zscore(Income), Zscore(Age)

d. Predictors: Zscore(Income), Zscore(Age), Zscore(Marital)

e. Predictors: Zscore(Income), Zscore(Age), Zscore(Marital), Zscore(Gender)

Appendix Table A.6 Stepwise Regression of Sociodemographic Variables

The segmentation model requires using demographic data as predictors. The above analysis gives us a comprehensive understanding of the data set. The following are results from the no-intercept linear (because of the standardization) regression with all remaining independent variables. The dependent variable is margin.

Model Summary

Model	R	R Square(a)	Adjusted R Square	Std. Error of the Estimate
1	.152(b)	.023	.023	.98376442

Appendix Table A.7 Summary of Regression Model

Coefficients[a,b]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	Zscore(Age)	-.088	.010	-.084	-8.761	.000
	Zscore(Gender)	.034	.009	.034	3.624	.000
	Zscore(Marital)	-.045	.012	-.041	-3.829	.000
	Zscore(WorkingWoman)	.013	.009	.015	1.524	.128
	Zscore(Income)	.131	.010	.126	12.961	.000
	Zscore(Children)	-.008	.010	-.008	-.841	.400
	Zscore(Adult)	.000	.010	.000	-.020	.984

a. Dependent Variable: Zscore(LnMargin)

b. Linear Regression through the Origin

Appendix Table A.8 Coefficients of Predictors

A.4 Cleaning Outliers and Cases with Missing Value(s)

We removed outliers (whose z-score value is equal to or bigger than 3) and cases having missing values. After data cleaning, there are 11,445 of the 20,921 cases are left. Because more than 50% of the data was removed, we need to standardize data from raw value again. The results are shown in the following tables.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age	11445	18.00	86.00	46.56	12.83
Gender	11445	.00	1.00	.72	.45
Marital	11445	.00	1.00	.79	.40
WorkingWoman	11445	.00	1.00	.47	.50
Children	11445	.00	1.00	.58	.49
Adult	11445	1.00	4.00	2.16	.70
Income	11445	1.00	9.00	6.29	2.04
Margin	11445	.82	140.22	14.74	12.46
Valid N (listwise)	11445				

Appendix Table A.9 Statistics of Raw Data

Descriptive Statistics

	N		Minimum		Maximum		Skewness	
	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
Zscore(Age)	11445		-2.23		3.08		.24	.02
Zscore(Gender)	11445		-1.60		.62		-.98	.02
Zscore(Marital)	11445		-1.97		.51		-1.46	.02
Zscore(WorkingW	11445		-.95		1.05		.10	.02
Zscore(Children)	11445		-1.16		.86		-.30	.02
Zscore(Adult)	11445		-1.66		2.63		.73	.02
Zscore(Income)	11445		-2.59		1.33		-.59	.02
Zscore(LnMargin)	11445		-3.27		3.22		-.22	.02
Valid N (listwise)	11445							

Appendix Table A.10 Statistics of Standardized Data

A.5 Random Sample

We randomly selected 1,500 customers from the remaining cases as the test data. Statistics of standardized data are similar to those of population because the mean closes to 0 and standard deviation is almost 1.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Zscore(Age)	1500	-2.23	3.08	.02	.99
Zscore(Gender)	1500	-1.60	.62	.00	1.00
Zscore(Marital)	1500	-1.97	.51	-.02	1.02
Zscore(WorkingWoman)	1500	-.95	1.05	.02	1.00
Zscore(Children)	1500	-1.16	.86	-.02	1.00
Zscore(Adult)	1500	-1.66	2.63	.01	1.01
Zscore(Income)	1500	-2.59	1.33	.00	1.00
Zscore(LnMargin)	1500	-3.24	2.99	-.01	.99
Valid N (listwise)	1500				

Appendix Table A.11 Descriptive Statistics of the Standardized Data

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age	1500	18.00	86.00	46.80	12.67
Gender	1500	.00	1.00	.72	.45
Marital	1500	.00	1.00	.79	.41
WorkingWoman	1500	.00	1.00	.48	.50
Children	1500	.00	1.00	.57	.50
Adult	1500	1.00	4.00	2.17	.71
Income	1500	1.00	9.00	6.28	2.05
Margin	1500	.84	117.22	14.49	11.98
Valid N (listwise)	1500				

Appendix Table A.12 Descriptive Statistics of the Raw Data

APPENDIX B DATA PREPARATION TWO

B.1 Introduction

The raw data is described in Appendix A. We started with the cleaned and coded data as input for the logistic regression model selection. We ran model selection before filtering out cases with missing values because removing variables first results in more cases without missing values. All purchased data are transformed using ln() to reduce skewness and to improve normality.

B.2 Model Selection

Sales and Margin are highly correlated with a Pearson correlation value of 0,822. So we removed Sales based on the order of entry in forward conditional logistic regression. ADBP and Purchase are highly correlated with a Pearson correlation value 0f 0.498. We kept Purchase but not ADBP because there was no missing value in Purchase. Running forward conditional regression with a stepwise entry probability of 0.01, the result is as follows.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	DSLP	-.773	.022	1274.632	1	.000	.462
	Constant	.464	.020	535.901	1	.000	1.591
Step 2[b]	DSLP	-.385	.025	236.902	1	.000	.680
	Purchase	1.202	.049	605.879	1	.000	3.327
	Constant	.659	.024	767.621	1	.000	1.934
Step 3[c]	Age	.165	.022	56.347	1	.000	1.180
	DSLP	-.380	.025	229.957	1	.000	.684
	Purchase	1.200	.049	600.949	1	.000	3.321
	Constant	.652	.024	745.292	1	.000	1.919
Step 4[d]	Age	.159	.022	51.788	1	.000	1.172
	DSLP	-.386	.025	236.625	1	.000	.680
	Margin	-.124	.021	34.365	1	.000	.883
	Purchase	1.198	.049	597.571	1	.000	3.315
	Constant	.652	.024	745.376	1	.000	1.920
Step 5[e]	Age	.141	.023	37.759	1	.000	1.151
	Children	-.065	.022	8.977	1	.003	.937
	DSLP	-.387	.025	237.917	1	.000	.679
	Margin	-.123	.021	33.650	1	.000	.884
	Purchase	1.192	.049	590.875	1	.000	3.294
	Constant	.651	.024	743.353	1	.000	1.918
Step 6[f]	Age	.139	.023	36.902	1	.000	1.149
	Children	-.074	.022	11.361	1	.001	.929
	Income	.058	.022	6.854	1	.009	1.060
	DSLP	-.385	.025	235.430	1	.000	.680
	Margin	-.128	.021	36.039	1	.000	.880
	Purchase	1.192	.049	591.125	1	.000	3.295
	Constant	.646	.024	725.165	1	.000	1.908

a. Variable(s) entered on step 1: DSLP.

b. Variable(s) entered on step 2: Purchase.

c. Variable(s) entered on step 3: Age.

d. Variable(s) entered on step 4: Margin.

e. Variable(s) entered on step 5: Children.

f. Variable(s) entered on step 6: Income.

Appendix Table B.1 Stepwise Logistic Regression

The Collinearity test shows that collinearity is not a concern (Tolerances close to 1, VIF, Eigenvalue and Condition Index are small).

Coefficients[a]

Model		Collinearity Statistics Tolerance	VIF
1	Zscore(Age)	.910	1.099
	Zscore(Children)	.893	1.120
	Zscore(Income)	.968	1.033
	Zscore(DSLP)	.794	1.259
	Zscore(Margin)	.986	1.014
	Zscore(Purchase)	.796	1.257

a. Dependent Variable: Target

Appendix Table B.2 Collinearity Statistics of Predictors

Collinearity Diagnostics[a]

Model		Eigenvalue	Condition Index	Variance Proportions (Constant)	Zscore (Age)	Zscore(Children)	Zscore(Income)	Zscore (DSLP)	Zscore (Margin)	Zscore(Purchase)
1	1	1.516	1.000	.00	.08	.08	.01	.19	.01	.19
	2	1.301	1.080	.00	.15	.20	.12	.09	.06	.07
	3	1.108	1.170	.40	.11	.01	.27	.00	.06	.01
	4	.981	1.243	.20	.01	.06	.00	.00	.71	.01
	5	.870	1.321	.39	.13	.00	.45	.00	.14	.00
	6	.678	1.495	.00	.51	.63	.15	.01	.02	.01
	7	.546	1.667	.00	.01	.01	.00	.71	.01	.71

a. Dependent Variable: Target

Appendix Table B.3 Collinearity Diagnostics of Predictors

B.3 Cleaning Cases and Random Sample

We removed the cases with missing values, leaving 12243 out of 20921. One thousand five hundred customers were randomly selected from the remaining cases as the test data.

Their statistics are shown in the following tables. Statistics of standardized data are similar to those of population because the mean closes to 0 and standard deviation is almost 1.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Target	1500	0	1	.62	.486
Zscore(Age)	1500	-2.03660	3.11582	.0533191	.92816686
Zscore(Children)	1500	-1.20312	.83111	-.0124199	1.00252605
Zscore(Income)	1500	-2.36244	1.35649	.1245967	.95733061
Zscore(DSLP)	1500	-1.00864	2.86896	.0118454	.99065838
Zscore(Margin)	1500	-1.74739	14.50910	.0524613	1.09690491
Zscore(Purchase)	1500	-.96707	7.21921	-.0480955	.81979505
Valid N (listwise)	1500				

Appendix Table B.4 Descriptive Statistics of the Standardized Data

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Target	1500	0	1	.62	.486
Age	1500	18	88	46.39	12.610
Children	1500	0	1	.59	.493
Income	1500	1	9	6.35	2.059
DSLP	1500	0	184	48.42	47.009
Margin	1500	-9.196667	210.8907	15.17050	14.850375230
Purchase	1500	0	44	4.94	4.406
Valid N (listwise)	1500				

Appendix Table B.5 Descriptive Statistics of the Raw Data

REFERENCES

Andrews, R. L., A. Ansari, et al. (2002). "Hierarchical Bayes Versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery." Journal of Marketing Research **39**(1): 87-98.

Arabie, P. and L. Hubert (1994). Cluster Analysis in Marketing Research. Advanced Methods of Marketing Research. R. P. Bagozzi. Cambridge, Balckwell Business**:** 160-189.

Back, T. (1996). Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms. New York, New York, Oxford University Press.

Bock, T. and M. Uncles (2002). "A Taxonomy of Differences between Consumers for Market Segmentation." International Journal of Research in Marketing **19**(3): 185-213.

Boulding, W., R. Staelin, et al. (2005). "A Customer Relationship Management Roadmap: What Is Known, Potential Pitfalls, and Where to Go." Journal of Marketing **69**: 155-166.

Branke, J., K. Deb, et al. (2004). Finding Knees in Multi-objective Optimization, Indian Institute of Technology Kanpur.

Brusco, M. J., J. D. Cradit, et al. (2002). "A Simulated Annealing Heuristic for a Bicriterion Partitioning Problem in Market Segmentation." Journal of Marketing Research **34**: 99-109.

Brusco, M. J., J. D. Cradit, et al. (2003). "Multicriterion Clusterwise Regression for Joint Segmentation Settings: An Application to Customer Value." Journal of Marketing Research **40**(2): 225-234.

Carroll, J. D. (1976). "Spatial, non-spatial and Hybrid Models for Scaling." Psychometrika **41**(4): 439-463.

Charnes, A. and W. W. Cooper (1961). "Management Models and Industrial Applications of Linear programming." Managment Science **15**(8): B365-B375.

Chesseman, P. (1996). Bayesian Classification (AutoClass): Theory and Results. Advances in Knowledge Discovery and Data Mining. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy. Cambridge, MA, AAAI/MIT Press**:** 153-180.

Claycamp, H. J. and W. F. Massy (1968). "A Theory of Market Segmentation." Journal of Marketing Research **5**(4): 388-94.

Coello, C. A. C., D. A. V. Veldhuizen, et al. (2002). Evolutionary Algorithms for Solving Multi-Objective Problems. New York, N.Y, Kluwer Academic Publishers.

Corne, D. W., N. R. Jerram, et al. (2001). PESA-II: Region-Based Selection in Evolutionary Multiobjective Optimization. Proceedings of the Genetic and

Evolutionary Computation Conference, San Francisco, California, Morgan Kaufmann Publishers.

DeSarbo, W. S. and D. Grisaffe (1998). "Combinatorial Optimization Approaches to Constrained Market Segmentation: An Application to Industrial Market Segmentation." Marketing Letters **9**(2): 115-134.

Dillon, W. R. and A. Kumar (1994). Latent Structure and Other Mixture Models in Marketing: An Integrative Survey and Overview. Advanced Methods for Marketing Research. R. P. Bagozzi, Blackwell Business**:** 353-388.

Dorigo, M., V. Maniezzo, et al. (1996). "The Ant System:Optimization by a Colony of Cooperating Agents." IEEE Transactions on Systems, Man, and Cybernetics–Part B **26**(1): 1-13.

E.Frank, R., W. F. Massey, et al. (1972). Market Segmentation. Englewood Cliffs, New Jersey, Prentice-Hall Inc.

Elliott, G. and W. Glynn (2000). Segmenting Industrial Buyers by Loyalty and Value. 16th IMP Conference, Bath.

Fraley, C. and A. E. Raftery (1998). "How Many Clusters? Which Clustering Methods? Answers Via Model-Based Cluster Analysis." The Computer Journal **41**(8): 578-588.

Frank, R. E. and P. E. Green (1968). "Numerical Taxonomy in Marketing Analysis: A Review Article." Journal of Marketing Research **5**(1): 83-94.

Frank, R. E., W. F. Massey, et al. (1972). Market Segmentation. Englewood Cliffs, New Jersey, Prentice-Hall Inc.

Frank, R. E. and W. F. Massy (1967). Effects of Short-Term Promotional Strategy in Selected Market Segments. Promotional Decision Using Methematic Models. P. J. Robinson. Boston, MA, Allyn and Bacon**:** 149-225.

Freitas, A. (2004). "A Critical Review of Multi-Objective Optimization in Data Mining: a Position Paper." SIGKDD Explorations **6**(2): 77-86.

Glover, F. W. and M. Laguna (2004). Tabu Search, Springer.

Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Reading, MA, Addison-Wesley Publishing Company.

Green, P. E. (1977). "A New Approach to Market Segmentation." Business Horizons **20**: 61-73.

Green, P. E., R. E. Frank, et al. (1967). "Cluster Analysis in Test Market Selection." Managment Science **13**(8): B387-400.

Green, P. E. and A. M. Krieger (1991). "Segmenting Markets with Conjoint Analysis." Journal of Marketing **55**(4): 20-31.

Grun, B. and F. Leisch (2006). "Fitting Finite Mixtures of Generalized Linear Regression in R." Computational Statistics & Data Analysis **In Press, Corrected Proof**.

Handl, J. and J. Knowles (2004). Evolutionary multiobjective clustering. Eighth International Conference on Parallel Problem Solving from Nature, Birmingham, UK.

Handl, J. and J. Knowles (2004). Multiobjective Clustering with Automatic Determination of the Number of Clusters. Manchester, UK, UMIST.

Handl, J. and J. Knowles (2004). Multiobjective Clustering with Automatic Determination of the Numbers of Clusters. Manchester, University of Manchester.

Hevner, A. R., S. T. March, et al. (2004). "Design Science in Information Systems Research." MIS Quarterly **28**(1): 75-105.

Hoch, S. J. (1994). Expterts and Models in Combination. The Marketing Information Revolution. R. C. Blattberg, R. Glazer and J. D. C. Little. Boston, Massachusetts, Harvard Business School Press**:** 253-269.

Hruschka, H. (1986). "Market Definition and Segmentation Using Fuzzy Clustering Methods." International Journal of Research in Marketing **3**: 117-134.

Ishibuchi, H. and Y. Shibata (2004). Mating Scheme for Controlling the Diversity-Convergence Balance for Multiobjective Optimization. 2004 Genetic and Evolutionary Computation Conference, Seattle, WA, USA.

Jain, A. K. and R. C. Dubes (1988). Algorithms for Clustering Data. EngleWood Clifs, NJ, Prentice Hall.

Jain, A. K., M. N. Murthy, et al. (1999). "Data Clustering: A review." ACM Computing Surveys **31**(3): 264-323.

Kiang, M. Y., M. Y. Hu, et al. (2006). "An Extended Self-organizing Map Network for Market Segmentation—a Telecommunication example." Decision Support Systems **42**(1): 36-47.

Kiang, M. Y. and A. Kumar (2001). "An Evaluation of Self-Organizing Map Networks as a Robust Alternative to Factor Analysis in Data Mining Applications." Information Systems Research **12**(2): 177-194.

Kirkpatrick, S., C. D. Gelatt Jr., et al. (1983). "Optimization by Simulated Annealing." Science **222**(4598): 671-680.

Kleinberg, J. (2002). An Impossibility Theorem for Clustering. Proceedings of the 15th Conference on Neural Information Processing Systems (NIPS). Vancouver, Canada.

Kleinberg, J., C. Papadimitriou, et al. (1998). Segmentation Problems. Proceedings of the Thirtieth Annual ACM Symposium on Thoery of Computing. Dallas, Texas, ACM Press**:** 473-482.

Kotler, P. (2003). Marketing Management (11th Edition), Pearson Education, Inc.

Krieger, A. M. and P. E. Green (1996). "Modifying Cluster-Based Segments to Enhance Agreement With an Exogenous Response Variable." Journal of Marketing Research **33**: 351-363.

Kuhn, H. W. and A. W. Tucker (1951). Nonlinear Programming. Processings of the Second Berkeley Symposium on Mathematical Statistics and Probability. J. Neyman. Berkeley, California, University of California Press**:** 481-492.

Leisch, F. (2004). "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R." Journal of Statistical Software **11**(8).

March, S. T. and G. Smith (1995). "Design and natural Science Research on Information Technology." Decision Support Systems **15**(4): 251-266.

Mattson, C. A., A. A. Mullur, et al. (2004). "Smart Pareto Filter: Obtaining A Minimal Representation of MultiObjective Design Space." Engineering Optimization **36**(6): 721-740.

McCann, J. M. (1974). "Market Segment Response to the Marketing Decision Variables." Journal of Marketing Research **11**(4): 399-412.

McDonald, M. and I. Dunbar (2004). Market Segmentation : How to Do It, How to Profit from It, Elsevier Butterworth-Heinemann.

Metropolis, N., A. W. Rosenbluth, et al. (1953). "Equation of State Calculations by Fast Computing Machines"." Journal of Chemical Physics **21**(6): 1087-1092.

Millgan, G. W. and M. C. Cooper (1988). "A Study of Standardization of Variables in Cluster Analysis." Journal of Classification **5**(2): 181-204.

Milligan, G. W. and M. C. Cooper (1988). "A Study of Standardization of Variables in Cluster Analysis." Journal of Classification **5**(2): 181-204.

Mooijaart, A. and P. G. M. Van Der Heijden (1992). "The EM Algorithm for Latent Class Analysis with Equality Constraints." Psychometrika **57**: 516-523.

Morwitz, V. G. and D. C. Schmittlein (1992). "Using Segmentation to Improve Sales Forecasts Based on Purchase Intent: Which Intenders Actually Busy"?" Journal of Marketing Research **29**: 391-405.

Moscato, P. (1999). Memetic Algorithm: a Short Introduction. New Ideas in Optimization. D. Corne, M. Dorigo, F. W. Gloveret al. Maindenhead, UK, McGraw-Hill Ltd., UK**:** 219-234.

Myers, J. (1996). Segmentation and Positioning for Strategic Marketing Decisions. Chicago, IL, USA, American Marketing Association.

Myers, J. H. (1996). Segmentation and Positioning for Strategic Marketing Decisions. Chicago, IL, American Marketing Association.

Osyczka, A. (1985). Multicriterion Optimization for Engineering Design. Design Optimization. J. S. Gero, Academic Press**:** 193-227.

Padmanabhan, B. and A. Tuzhilin (2003). "On the Use of Optimization for Data Mining: Theoretical Interactions and eCRM Opportunities." Managment Science **49**(10): 1327-1343.

Pareto, V. (1896). Cours D'Economi Politique. Rouge, Lausanne.

Park, Y.-J. and M.-S. Song (1998). A Genetic Algorithm for Clustering Problems. Proceedings of the Third Annual Conference on Genetic Programming. Madison, WI, Morgan Kaufmann**:** 568-575.

Payne, A. and P. Frow (2005). "A Strategic Framework for Customer Relationship Management." Journal of Marketing **69**: 167-176.

Pepper, D. and M. Rogers (2004). Management Customer Relationships: A strategic Framework. Hibijenm New Jersey, John Wiley & Sons, Inc.

Peppers, D., M. Rogers, et al. (1999). "Is Your Company Ready for One-to-One Marketing?" Harvard Business Review **77**: 151-160.

Phelps, S. P. and M. Koksalan (2003). "An Interactive Evolutionary Metaheuristic for Multiobjective Combinatorial Optimization." Management Science **49**(12): 1726-1738.

Punj, G. and D. W. Stewart (1983). "Cluster Analysis in Marketing Research: Review and Suggestion for Application." Journal of Marketing Research **20**: 134-18.

R Development Core Team. (2007). "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing Retrieved May 16, 2007, 2007, from http://www.R-project.org.

Ramaswamy, V., R. Chatterjee, et al. (1996). "Joint Segmentation on Distinct Interdependent Bases with Categorical Data." Journal of Marketing Research **33**(3): 337-350.

Rigby, D. (2005) "Management Tools 2005." New York, Bain & Company **Volume**, DOI:

Rosenberg, R. S. (1967). Simulation of genetic populations with biochemical properties. Ann Arbor, University of Michigan.

Rubin, M. (1997). "Creating Customer-Oriented Companies." Prism **4**(4): 5-27.

Schaffer, D. (1984). Multiple Objective Optimization with Vector Evaluated Genetic Algorithms. Nashville, Tennessee, Vanderbilt University.

Skinne, H. A. (1978). "Differentiating the Contribution of Elevation, Scatter and Shape in Profile Similarity." Educational and Psychological Measurement **38**: 297-308.

Skinner, H. (1978). "Differentiating the Contribution of Elevation, Scatter and Shape in Profile Similarity." Educational and Psychological Measurement **38**: 297-308.

Smith, W. R. (1956). "Product Differentiation and Market Segmentation as Alternative Marketing Strategies." Journal of Marketing(21): 3-8.

Spath, H. (1979). "Clusterwise Linear Regression." Computing **22**: 367-373.

Spath, H. (1982). "A Fast Algorithm for Clusterwise Linear Regression." Computing **29**: 175-181.

Tibshirani, R., G. Walther, et al. (2000). Estimating the Number of Clusters in a Dataset vai the Gap Statistic, Dept. of Statistics, Stanford University.

Titterington, D. M. (1990). "Some Recent Research in the Analysis of Mixture Distributions." Statistics **4**: 619-641.

Urbanek, S. (2003). Rserve -- A Fast Way to Provide R Functionality to Applications. 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, Austria, Achim Zeileis.

Urbanek, S. (2006). "About JRI." Online Document Retrieved May 16, 2007, 2007, from http://rosuda.org/JRI/.

Vargo, S. L. and R. F. Lusch (2004). "Evolving to a New Dominant Logic for Marketing." Journal of Marketing **68**: 1-17.

Vriens, M., M. Wedel, et al. (1996). "Metric Conjoint Segmentation Methods: A Monte Carlo Comparison." Journal of Marketing Research **33**(1): 73-85.

Wedel, M. and W. S. DeSarbo (1994). A Review of Latent Class Regression Models and Their Applications. Advanced Methods for Marketing Research. R. P. Bagozze, Blackwell Business**:** 353-338.

Wedel, M. and W. Kamakura (2000). Market Segmentation: Conceptual and Methodological Foundations. Norwell, Massachusetts, Kluwer Academic Publishers.

Weinstein, A. (1993). Market Selection in Technology-Based Industry: Insights from Executives. American Marketing Association Winter Educators' Conference Proceedings, Newport Beach, California.

Wind, Y. (1978). "Issues and Advances in Segmentation Research." Journal of Marketing Research **15**: 317-37.
Winer, R. S. (2001). "A Framework for Customer Relationship Management." California Management Review **43**(4): 89-105.
Zhong, S. and J. Ghosh (2003). "A Unified Framework for Model-based Clustering." Journal of Machine Learning Research **4**: 1001-1037.