CLUSTERWISE REGRESSION AND MARKET SEGMENTATION

-developments and applications-

Promotoren: dr. ir. M.T.G. Meulenberg,
hoogleraar in de Marktkunde en het Marktonderzoek

dr. P.S.H. Leeflang,
hoogleraar in de Marktkunde en het Marktonderzoek
aan de Rijksuniversiteit Groningen

Co-promotor: dr. ir. J.E.B.M. Steenkamp,
universitair docent in de Marktkunde en het Marktonderzoek

STELLINGEN

1. Voor het identificeren van segmenten van voldoende grootte, stabiliteit, reactiviteit en toegankelijkheid, en voor het genereren van aangrijpingspunten voor het ontwikkelen van marketingstrategieën, is "benefit"-segmentatie een van de meest effectieve segmentatiebases die ter beschikking staan.

   Dit proefschrift.

2. De "clusterwise regression"-methoden die in dit proefschrift zijn ontwikkeld en toegepast voor marketing van tastbare consumentenprodukten, kunnen evenzeer waardevolle bijdragen leveren aan segmentatieproblemen in vakgebieden als direct marketing en industriële marketing, alsmede de marketing van diensten.

   Dit proefschrift.

3. Ten behoeve van de ontwikkeling en toepassing van latente-klasse-regressiemodellen in segmentatie-onderzoek verdient de validiteit van de gebruikte asymptotische significantietoetsen voor de coefficiënten nader onderzoek.

   Dit proefschrift.

4. Het fundamentele verschil tussen de "fuzzy" clusteringsmethoden gebaseerd op de "fuzzy-set"-theorie, en die gebaseerd op latente-klasse-modellen is dat de eerste ervan uitgaan dat objecten partieel tot meerdere clusters behoren, hetgeen leidt tot een "fuzzy" classificatie, terwijl de tweede veronderstellen dat objecten tot één cluster behoren, zodat de "fuzzy" classificatie het gevolg is van onvoldoende informatie in de data om objecten eenduidig toe te wijzen.

   Dit proefschrift.

5. Clusteringsmethoden die overlappende clusters identificeren in een regressiecontext, zijn discutabel aangezien het regressiemodel additief is tussen twee overlappende clusters en derhalve objecten die tot beide clusters

Errata

Clusterwise regression and market segmentation
-developments and applications-
Michel Wedel

Page 157, heading of Table 9.3:
"FCRG   FCR", in columns 6 and 7 should read: "FCR   GFCR".

Page 160, Table 9.5 bottom part (Data set 2), Column 3:
"0.27, 0.24, 0.25, 0.23, 0.25" should be:
"0.45, 0.04, 0.33, 0.04, 0.66".

Page 163, first line under Figure 9.2, "From the plot" should be: "From the plot of", and "also be justified", in the second line under Figure 9.2 should read "also justified".

Page 207, line 6, "Karate championships and gained" should be: "Karate and gained".

MICHEL WEDEL

CLUSTERWISE REGRESSION AND MARKET SEGMENTATION

-developments and applications-

Proefschrift
ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen,
op gezag van de rector magnificus,
dr. H.C. van der Plas,
in het openbaar te verdedigen
op vrijdag 7 december 1990
des namiddags te vier uur in de aula
van de Landbouwuniversiteit te Wageningen

Voor B.

ACKNOWLEDGMENTS

The studies for this thesis were carried out at the TNO-CIVO Toxicology and Nutrition Institute, Zeist, Netherlands, and were financially supported by the Dutch Ministry of Agriculture, Fisheries and Nature Management.

Looking back on the five years it took to complete this work, there are many of people whom I wish to thank (in chronological order) for their contributions, to ideas, content and realization.

I am grateful to Professor Ruud J.J. Hermus, director of the Institute, for his devising and the project proposal that set me to work on the development of the methods that are the subject of this thesis.

Dr. Theo Ockhuizen, head of the Department of Human Nutrition of the Institute, has provided me with the necessary organizational conditions. I have highly appreciated his support.

Professor Matthieu T.G. Meulenberg, Wageningen Agricultural University has contributed to this work through the in-depth discussions on various topics. The reader finds his emphasis on coherence reflected in this work.

Professor Peter S.H. Leeflang of the University of Groningen has urged me to write an integral text, instead of compiling a mix of published papers and unpublished material. I am very grateful for the amount of time he has spent on reading and discussing the manuscript in great detail, especially with respect to the mathematical aspects.

Dr Jan-Benedict E.M. Steenkamp has contributed enormously. He has provided me with the data for the empirical applications in Chapters 8 and 9, and has been the co-author of a number of papers on which this thesis is based. We have worked together in a pleasant and stimulating way.

Ing. Cor Kistemaker, my colleague at the Department of Human Nutrition, has written the computer programs for the methods developed. He co-authored a paper incorporated in Chapter 7. I appreciate his contributions very much.

Many other colleagues at the Institute have contributed to this thesis. I am indebted to Dirk G. van der Heij for editorial work and to Evelien J.M. Aarnink and Pita M. van der Es for assistance in data processing, to Lyda Punt, Tineke W.G.M. van der Vaart and Henk J. de Vetter for assistance in typing, to Martin van de Vaart for making the drawings, and to Michiel R.H. Löwik and Karin F.A.M. Hulshof for advice.

Financial support by the LEB-fund and the Commodity Board for live stock and meat for publication of this thesis is gratefully acknowledged.

Finally, I wish to let the reader know that I could never have finished this work without the moral support and loving patience of my wife Hennie.

CONTENTS

1 INTRODUCTION

Modern marketing in industrialized countries cannot do without segmentation of the market of its potential customers. Goods can no longer be produced and sold without understanding consumer needs, and recognizing the heterogeneity of consumers' demand for products. Until the 20th century, marketing predominantly involved a strategy of customized marketing and merchandising, each customer being served individually according to her or his needs, within the constraints of limited supply alternatives of producers and disposable income of the average consumer.

In the beginning of this century, however, industrial development in various sectors of the economy induced strategies of mass marketing, and consumers were all offered the same, undifferentiated products. Due to mass production, products did not match the needs of all consumers optimally. But the flexibility of the production process increased, and the increasing affluence of consumers contributed to diversity in demand. Firms that identified the specific needs of groups of consumers could develop the right offer for each of these submarkets and obtained a competitive advantage.

The concept of market segmentation emerged, and catering to segments became an important aspect of competition between firms. Today's companies are increasingly focusing on mini-markets with different needs and different behavior towards distribution and communication channels. Segmentation has been considered one of the most fundamental concepts of modern marketing, and has spread beyond market research into areas such as food research and social research. The importance of the segmentation concept is reflected in the vast amount of segmentation studies conducted by firms and market research agencies, and the impressive amount of literature published on this topic in scientific journals.

Market segmentation is also the topic of this work, the main thrust being segmentation of consumer markets. The terms 'consumer segmentation', 'market segmentation' or 'segmentation' will be used interchangeably. Industrial segmentation or segmentation of the demand of services will not be considered.

The present work consists of two parts. Part 1 (Chapters 2 to 6) reviews the literature on segmentation and sets the stage for Part 2 (Chapters 7 to 11), in which a set of new methods for consumer segmentation are developed and applied.

Chapter 2 discusses the segmentation concept, marketing strategies, and research. A common understanding of the concept of consumer segmentation is essential as a basis for research on this topic, and its implications for marketing strategy. Once segments have been identified, firms may develop marketing strategies in which groups of consumers with different needs are targeted with a different marketing mix. Segmentation research provides management with information required for the development of marketing strategy. Two major streams of segmentation research are identified on the basis of their theoretical foundation, which can be based on either microeconomic theories or behavioral science. Both the methods and the variables used as a basis for segmentation differ according to these two streams.

Chapter 3 deals with variables that may be used as a basis for segmentation. Alternative bases are described and evaluated according to their effectiveness for segmentation.

The statistical methods that are available for segmentation research are described in Chapter 4. Segmentation research is a grouping task, for which a large variety of methods have been used. A major distinction is made between a priori and post hoc methods, which define the number of segments and their boundaries before, and after data collection respectively.

In Chapters 5 and 6, the approaches to segmentation developed within the two major streams of segmentation research are reviewed. In Chapter 6 focus is upon benefit segmentation, and the discussion sets the stage for the development of the new segmentation procedures in Part 2.

In Part 2 (Chapters 7 to 10) a set of new segmentation techniques are developed, called clusterwise regression methods, which combine prediction with classification. Chapter 7 describes a method that yields non-overlapping segments; the method is applied to the analysis of preferences for meat (A part of this chapter has been published before, Wedel and Kistemaker 1989). In Chapter 8 a method is developed that allows of fuzzy segmentation, in which consumers may belong to more than one segment (A part of this chapter was published, Wedel and

Steenkamp 1989). Two applications are provided: one entails the analysis of preferences for meat products, the second one entails an analysis of consumer attitudes towards outlets selling meat. Chapter 9 describes a method that incorporates simultaneous benefit segmentation (grouping consumers) and market structuring (grouping products). The method is applied to the analysis of preferences for butter and margarine brands.

The performance of the methods is critically evaluated using cross-validation procedures, and Monte Carlo studies on artificial data. The methods are compared with other segmentation methods described in the literature.

Finally, in Chapter 10, the implications of the analyses for a number of relevant fields of marketing research (e.g. food marketing) are discussed. Both the strong and weak points of the methods are critically evaluated, and suggestions for further developments and applications are made.

PART 1

AN OVERVIEW OF THE SEGMENTATION LITERATURE

2 MARKET SEGMENTATION

2.1 The concept

Whereas the theory of perfect competition assumes homogeneity among both the demand and supply sides of the market, the theory of imperfect competition (Robinson 1938) deals with diversified markets. Variations in production methods and resources, the increasing affluence of consumers, and variation in producers' estimates of the prospective consumer needs contributed to this diversification (Smith 1956).

Market segmentation has become a necessary perspective of the market for a business firm to provide an optimum match between the increased flexibility of production and the needs of homogeneous groups of consumers. It has been a central concept in both marketing theory and marketing practice since its introduction by Smith in 1956. Smith recognized the existence of heterogeneity in the demand of goods and services, based on the economic theory of imperfect competition. Smith defined: "Market segmentation involves viewing a heterogeneous market as a number of smaller homogeneous markets, in response to differing preferences, attributable to the desires of consumers for more precise satisfaction of their varying wants".

It was essentially Smith's (1956) work that prompted researchers to develop techniques for partitioning the market and describing the submarkets. The segmentation concept provided a basis for identifying the data needed for the selection and implementation of marketing strategies. Segmentation was extended, both theoretically and practically, to virtually all strategic and tactic marketing plans and programs.

However, there has been a lack of precision of the use of the term "market segmentation" in the marketing literature (Dickson and Ginter 1987). Three different views can be distinguished.

First, market segmentation is viewed by some authors as the recognition of the existence of heterogeneity in the market and the *development of a strategy* to cater to more homogeneous submarkets (Smith 1956, Engel, Fiorillo and Cayley 1972, Frank Massy and Wind 1972, Wilkie and Cohen 1977, Bass, Tigert and Londsdale 1968, and Winter 1979, 1984).

Second, opposing to the above view, is the use of the term "segmentation" to refer to *activities of identifying* homogeneous consumer groups (Johnson 1971, Mahajan and Jain 1978, Hayley 1968, and Kotler 1988).

Third, Dickson and Ginter (1987) defined market segmentation as a *state of the market* in which heterogeneity in demand exists, which allows of the identification of market segments. In the views expressed by Dickson and Ginter (1987) and Dickson (1982), segmentation does not entail a grouping of people, products or usage situations, but of demand functions.

In addition to these differing views of segmentation, some authors question whether natural groups or segments exist at all (Johnson 1971, Shepard and Arabie 1979, Arabie et al. 1981, Hagerty 1985). They argue that it is only convenient to approximate the heterogeneity of the market by clustering respondents into segments. The segmentation concept only imposes a possible partitioning on the market, which may at best provide an acceptable approximation to the real condition of market heterogeneity. Based upon these considerations we propose the following conceptualization of (consumer) market segmentation.

Segmentation is a theoretical marketing concept partitioning a market with heterogeneous demand into submarkets with homogeneous demand, with the purpose of a more precise adjustment of brands, products, or services to consumer needs, to determine the potentially most profitable allocation of marketing efforts.

The accuracy of the firm's identification of market segments contributes to its competitive advantage, because it is a prerequisite to the development of market segmentation strategies by a firm's management. A number of conditions affect the effectiveness and profitability of the marketing strategy (e.g. Frank et al. 1972, Baker 1988, Loudon and Della Bitta 1984, Kotler 1988):

- *Identifiability*. It should be possible to identify segments, and obtain relevant information on their consumers' characteristics. The greater the differences in prospect buying behavior between segments, the more justification for segmentation strategies.

- *Substantiality.* Segments should be large enough to warrant the profitability of a targeted market program. Both the number of customers and their purchasing power are important.
- *Responsiveness.* Segments should respond uniquely to marketing efforts targeted at them. Preferably, segments should constitute a gap in the market.
- *Accessibility.* It should be feasible to reach segments through promotional or distributional efforts.
- *Stability.* Segment membership, accessibility, size and responsiveness should be stable for a certain period of time at least to allow of predictions from the time of study to the time of implementation of the marketing strategy.
- *Actionability.* Segmentation studies should provide clues for strategic decisions on the effective specification of marketing instruments.

2.2 **Marketing strategies**

Firms may perceive different degrees of segmentation in the market, from completely aggregate to completely disaggregate. On the basis of information on the degree of segmentation, they may employ undifferentiated (or mass) marketing (ignoring segment differences and offering one brand to the entire market), differentiated marketing (operating in different segments with different brands and marketing mixes in each segment), or target marketing (concentrating on one segment) (Kotler 1988). Figure 2.1 depicts these strategies.

When a firm uses a target marketing strategy, it might either target the segment with one product (segment concentration) or serve many needs of a particular segment (market specialization). When a firm employs a differentiated marketing strategy, three possible market coverage patterns can be distinguished: selective specialization (the firm selects a number of segments each of which matches the company's objectives, and which are uniquely targeted), product specialization (the firm produces one product, which is sold to a number of segments), and full coverage (the firm serves all segments with all products they might need) (Kotler 1988).

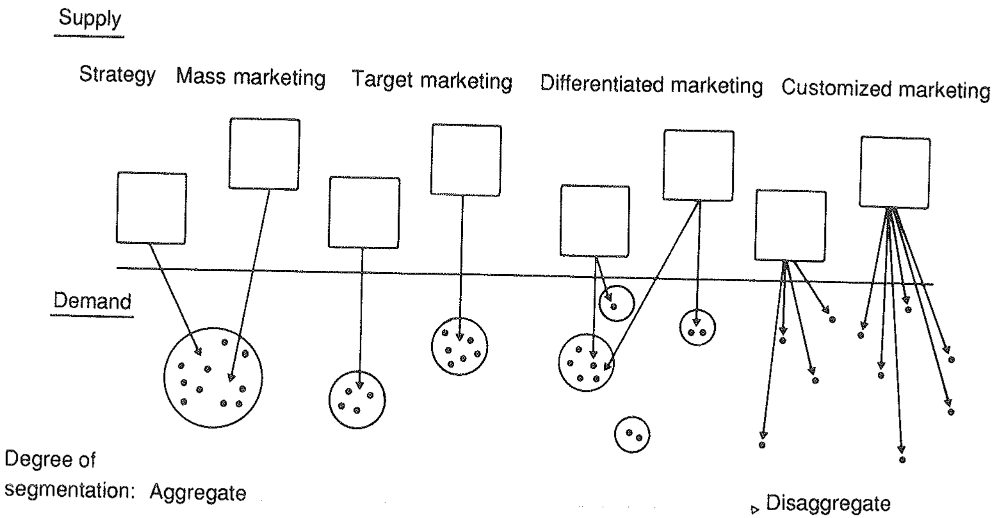

Figure 2.1 Market segmentation and market segmentation strategies. (The degree of segmentation increases from left to right, the squares denote the enterprises on the supply side, the dots and circles denote consumers and segments respectively on the demand side, the arrows denote the offers provided)

Product differentiation has been described both as a strategy opposing market segmentation (Kotrba 1966), and as a means of implementing market segmentation. The misunderstanding on this issue was clarified by Dickson and Ginter (1987), who defined product differentiation as the marketplace condition in which not all products are perceived by consumers as being equal. A product differentiation strategy was defined as the deliberate alteration of consumers' perceptions of product attributes. This can be achieved by product development, product modification, pricing, or communication (Curry and Menasco 1983, Leeflang 1987).

Alternatively, firms may attempt to increase the importance consumers attach to an attribute on which the product in question has a competitive advantage. This (product-oriented) strategy was called demand function modification by Dickson and Ginter (1987). Demand function modification can be achieved by promotional activities (Curry and Menasco 1983, Leeflang 1987).

The development of marketing strategy is thus critically dependent upon the current market condition and the accuracy with which it is perceived by the firm's management. Segmentation research plays an important role herein.

2.3 Segmentation research

Frank, Massy and Wind (1972) divided research on market segmentation into the decision-oriented school and the behavior-oriented school. The major difference between these two is their theoretical orientation: the first derives from microeconomic theory, whereas the second is founded upon theories originating from the behavioral sciences. In decision-oriented segmentation research, the question of which and how existing segments respond to marketing instruments is the issue, rather than the question why segments exist, or how they can be identified. These are the issues addressed by the behavioral school of segmentation research. It focuses on approaches for the identification of segments and searches for variables that contribute to the theory of why the differences in behavior between segments occur.

Wilkie and Cohen (1977) expanded upon the scheme of Frank, Massy and Wind (1972) and discriminated between the correlational stream and the product-instrumentality stream. According to their definition the correlational stream searches for general consumer characteristics that are correlated with actual behavior, be it level of purchase or purchase response to marketing variables. The definition of the correlational stream is thus somewhat wider than that of the decision- oriented stream of Frank, Massy and Wind (1972). The product-instrumentality stream is defined by Wilkie and Cohen (1977) to be founded upon behavioral science theories.

We will focus on microeconomically oriented (choice-based) segmentation in Chapter 5, and behavioral science-oriented segmentation in Chapter 6. Although the approaches towards segmentation have blended (e.g. Rao and Winter 1978, Hauser and Urban 1977, Winter 1979, Currim 1981), this categorization appears a useful way to separate two research traditions. The chapter on the microeconomic school includes the correlational (Wilkie and Cohen 1977) and the decision-oriented approach (Frank, Massy and Wind 1972). The research in this school is outcome-oriented, as it aims at the description of differences in choice behavior between segments, either at the cross-sectional level or with respect to the response to marketing variables. The behavioral school can be called process-oriented: its main thrust is understanding segment differences in consumer behavior on the basis of behavioral science theory.

The differences between the two research traditions pertain not only to the theoretical underpinnings, but also to the bases and methods that are used to identify segments. Bases and methods for segmentation will be described in Chapters 3 and 4 respectively.

A segmentation basis is defined as the characteristic or groups of characteristics of consumers, used to assign consumers to segments.

Frank, Massy and Wind (1972) classified segmentation bases into general (independent of any product or service and the particular circumstances faced by the consumer) or situation-specific (related to both the consumer and the product, service and/or particular circumstances). The latter bases have been referred to alternatively as behavioristic (Baker 1988) or product-instrumental (Wilkie and Cohen 1977) we will use the term product-specific to avoid confusion with usage situations. Furthermore, Frank et al. (1972) classified bases according to whether they can be measured objectively (observable bases), or have to be inferred (unobservable bases). This classification system results in four classes (Figure 3.1).

Figure 3.1. Classification of segmentation bases[1]

	General	Product-specific
Observable	Cultural, geographic, demographic and socio-economic variables. (3.2)	User status, usage frequency, brand loyalty, store loyalty and patronage, usage situation. (3.3)
Unobservable	Psychographics: personality and life-style (3.4)	Psychographics, benefits, perceptions, attitudes, preferences and intentions (3.5)

[1] The numbers in parentheses refer to sections containing a treatment of the bases in question.

This chapter provides an overview of the bases in each of the four resulting classes. (See, for example, also Frank, Massy and Wind 1972, Boyd and Massy 1972, Wilkie and Cohen 1977, Loudon and Della Bitta 1984, Baker 1988, Leeflang 1987, and Kotler 1988)[1)]. Each of the bases will be discussed in terms of the criteria for effective segmentation, described in Section 2.1.

Before we turn to a description of the four classes of bases we will elaborate on the bases that have been put forward as the normative ideal for segmentation research (these bases could alternatively be classified as unobservable product-specific bases).

3.1 The normative ideal basis

The microeconomic school of segmentation research has concerned itself with the normative ideal basis for segmentation. Inherent in its study of consumer demand, this school has initially suggested normative ideal bases in relation to consumers' demand response functions (Frank and Massy 1965, Claycamp and Massy 1968). As opposed to normative segmentation *bases*, some have proposed normative segmentation *methods* (e.g. Frank, Massy and Wind 1972, Tollefson and Lessig 1978) that result in maximum profit for a firm. The segmentation resulting from profit maximization is specific to the firm in question (often through a cost function), contrary to segments derived from normative ideal bases which are idiosyncratic to the demand side only.

Whereas Frank and Massy (1965), Dhalla and Mahatoo (1976), and Sexton (1974) articulated elasticities to be the normative ideal basis for segmentation, Claycamp and Massy (1968) suggested the individual marginal responses to marketing variables (both of these measures were only used with respect to price and promotion). In their attempt to alleviate the apparent confusion, Dickson and Ginter (1987) claimed that the demand function itself is the normative ideal basis for segmentation. However, the question remains of how demand functions should be characterized in practice.

This question was empirically investigated by Tollefson and Lessig (1978) and Elrod and Winer (1982). Their analyses of (simulated respectively empirical) data of individual consumers response to marketing variables revealed that the values of the marketing variables needed for each consumer to optimize his demand function (referred to as optimal disaggregate allocation) performed substantially better than 1. elasticities, 2. marginal response, and 3. demand function coefficients, with respect to the resulting profits.

Blozan and Prabhaker (1984), however, showed that the evidence presented by Tollefson and Lessig (1978) is not valid. They demonstrated that, in general, reliance on a local measure of demand (elasticities, coefficients, etc.) is not obviously unsatisfactory. The suboptimality of elasticities found by Tollefson and Lessig (1978) and Elrod and Winer (1982) may be due to either the ineffectiveness of the elasticity measure, or the clustering procedures applied. These fail to detect the optimal solution because the criterion of homogeneity of elasticities upon which they rely is not necessarily the optimal criterion for grouping elasticities, except in the case of linear demand functions. Consequently, general theoretical arguments supporting the use of a local measure of demand for the aggregation of consumers into segments are not available, but in general the disaggregate allocation rule is expected to perform well (Blozan and Prabhaker 1984).

Although criteria such as elasticities or demand function coefficients are considered ideal from a normative point of view, in practice there are few who have used it as a basis for segmentation, (e.g Elrod and Winer 1982). The reason is that the estimation of the criteria from individual consumers' responses to marketing variables, and thereby the identification of segments, is difficult. (Besides, the elasticity with respect to e.g. distribution are hard to estimate.)

Because of the problems with the implementation of the normative ideal in practice, the search for other bases has flourished. The normative ideal basis has instead been used to assess the effectiveness of segments identified from other bases, in terms of the differences in their response to marketing variables.

3.2 Observable general bases

A number of bases fall into this category:

1. Cultural variables, such as race, nationality and religious groupings (Frank et al. 1972).
2. Geographic variables, such as region, population density, climate (Frank et al. 1972), neighborhood (Baker 1988), and geographic mobility (Loudon and Della Bita 1984).
3. Demographic variables, such as age, sex, marital status, family size, stage in life cycle (Frank et al. 1972).

4. Socioeconomic variables, such as income, occupation, education (Frank et al. 1972), expenditures (Loudon and Della Bita 1984), socioeconomic class (Frank et al. 1972), and time and money (Leeflang 1987).

These segmentation bases are easy to collect, reliable, and generally stable. The results are easy to communicate and resulting strategies easy to implement. However, these bases are less actionable, than psychographics, benefits etc., in the sense that they provide clues on how to communicate to the segments. Segments based upon this class of variables are readily accessible because of the wide availability of media profiles for most of the bases mentioned (Frank, Massy and Wind 1972).

There has been much research into the effectiveness of observable general bases as predictors of purchase behavior (in relation to the responsiveness criterion for effective segmentation). Some differences in purchase behavior among these type of segments have been found (see e.g. Frank et al. 1972), but in many studies the degrees of association of variables such as family size, education and occupation, with purchase behavior were found to be weak (Frank 1972). Both Frank (1972) and McCann (1974) investigated the effectiveness of this class of segmentation bases in terms of their response to marketing mix variables (price and deals). The studies concerned food products. Although some differences in elasticities of the marketing variables according to education, income, employment status, expenditures, age and household size were found, the lack of strongly significant findings support the conclusion that these variables are not particularly effective segmentation bases.

Yet, these findings have not resulted in a complete discard of the general and observable bases. They continue to be used as an element in segmentation approaches, often as segment descriptors, because of their appeal with respect to identifiability, measurability, accessibility and stability.

3.3 Observable product-specific bases

The bases in this class comprise both variables related to purchase behavior and usage situational variables:

1. User status; consumers are classified as users or nonusers of a product class, a product or a brand (Boyd and Massy 1972, Frank et al. 1972).
2. Usage frequency; consumers are classified as e.g. heavy, moderate or light users (Twedt 1967).
3. Brand loyalty; consumers are classified as e.g. hard-core loyals, soft loyals, shifting loyals or switchers (Boyd and Massy 1972).
4. Store loyalty (Loudon and Della Bita 1984, Frank, Massy and Wind 1972).
5. Store patronage; this divides the market into groups on the basis of the type of shoppers (Frank, Massy and Wind 1972).
6. Participation in the adoption process; for example, innovators, early adopters, early majority, late majority, laggards (Frank, Massy and Wind 1972, Leeflang 1987).
7. Usage situational variables; these are factors, other than consumer and product, that influence choice behavior, such as physical surroundings, temporal perspective, task definition, social surroundings, antecedent states (Loudon and Della Bita 1984, Dickson 1982, Belk 1975).

3.3.1 *Purchase behavior*

The segmentation variables based upon purchase behavior are often used in the microeconomic stream. Data on these segmentation variables are relatively easy to collect. Accessibility of the segments has to be established by relating them to general consumer descriptors. These relations are however weak (Frank 1972, Frank, Massy and Wind 1972).

The differences in response to marketing variables (price and dealing) of segments derived from this class of segmentation bases was investigated by Frank (1967, 1972), Massy and Frank (1965), Sexton (1974), and McCann (1974). No differences in response to marketing variables between segments with a different degree of brand loyalty were found in these studies, although loyalty was found to be a stable concept. Weak to moderately significant differences in elasticities were found according to usage frequency, store loyalty, store patronage, deal-proneness, and innovativeness.

Consequently, there is some evidence that purchase behavior variables can be used as a basis for segmentation. Their association with consumer response to marketing variables is somewhat stronger than that of general consumer descriptors. Besides, these segments are appealing because of their measurability and identifiability. Accessibility may pose a problem because of the low degree of association with general consumer descriptors.

3.3.2 *Usage situations*

Dickson (1982) provided a general theoretical framework for usage situation as a segmentation basis. Behavior originates from the interaction between a consumer and his environment, and the demand that results can be conceptualized as an aggregation of demand in different situations. When the situational demand functions are substantially heterogeneous, situational segmentation is theoretically viable.

Situation-based variables are directly measurable, and the corresponding segments are thus identifiable. The degree to which segments are substantial depends on the frequency of occurrence of the usage situation. The segments are accessible because consumers can sometimes be reached in the usage situations, or alternatively because the consumers can be identified in specific usage situations by general descriptors (Dickson 1982). Stout et al. (1977) demonstrated that usage situational segments may be quite different from one another with respect to these consumer descriptors. The responsiveness of usage situational segments was investigated by Belk (1975), Stout et al. (1977), and Miller and Ginter (1979). Perceptions of product attributes, their importances, buying intentions and purchase frequency and volume were all found to differ significantly across usage situations.

Consequently, the explicit consideration of situational contexts contributes to the explanation of consumer behavior and appears to be a promising direction in segmentation research when used in addition to (rather than instead of) consumer segmentation.

3.4 Unobservable general bases

The segmentation variables within this class fall into two major types:
1. Personality traits;
2. Life-style.

3.4.1 *Personality*

Frank, Massy and Wind (1972) defined personality as the configuration of individual psychological characteristics that differentiates a person from others. Edward's personal preference schedule (EPPS, cf. Frank et al. 1972) is the most frequently used tool for measuring general aspects of personality. Applications include those of Evans (1959), Frank (1972) and Koponen (1960). The EPPS appraises personality traits such as achievement, order, dominance, and endurance. Riesmann's measure of social character (cf. Frank et al. 1972) allows for segmentation into inner-directed (turning to their own inner values for guidance of their actions), other-directed (depending on the people around them) and tradition-directed (oriented in a traditional way).

A methodological problem in the application of personality scales in market segmentation has been the lack of attention to the level of psychological functioning represented, in relation to the purchase behavior being studied (Wells 1975, Wilkie and Cohen 1977). General personality measures more likely show a relationship with patterns of behavior (such as innovativeness) than with behavior with respect to a single product or brand (Verhallen and Pieters 1984). Frank (1972), Frank, Massy and Wind (1972) and Wells (1975) summarize a number of studies in which a relationship between personality and purchase behavior or purchase related variables were assessed. The personality scales had at best a low degree of association with purchase behavior.

3.4.2 *Life-style*

The concept of life-style was introduced into marketing by Lazer (1963). Its aim is to draw a recognizable picture of consumers, based on activities, interests and opinions (AIOs). Life-style has evolved from a blend of the traditions of personality inventories and motivation research (see Wells 1975). Plummer (1974) defined the major life-style dimensions:

1. Activities : work, hobbies, social events, vacation, entertainment, club membership, community, shopping, sports;
2. Interests : family, home, job, communication, recreation, fashion, food, media, achievements;
3. Opinions : themselves, social issues, politics, business, economics, education, products, culture, future.

Some researchers also include values in life-style scales. Mitchell (1983) proposed a values and life-style typology, representing the degree to which consumers are stereotypical sustainers or survivors. The recent analyses of Lastovicka et al. (1990), however, revealed negative evidence for the validity of the system. Frank, Massy and Wind (1972) and Wells (1975) summarized a large number of studies in which relationships between life-style and purchase behavior were assessed. Wells concluded that the predictive validity of life- style can be substantially higher than that of general observable segmentation bases. Life-style scales should at least have some hypothetical relationship with the behavior being studied. Hereby, the usefulness of general life-style measures for market segmentation is implicitly questioned. Amongst others, Ziff (1971), Wells (1975), and Dickson (1982) posit that general life-style segmentation suffers the same defects as general personality scales in that life-style is likely to identify basic attitudes that influence extensive behavioral patterns in many situations, rather than any specific behavior towards a product or brand. Moreover, because individuals play different roles in different situations, life-style can not be used without consideration of the situational context (Dickson 1982). Frank, Massy and Wind (1972) point to the lack of theory of life-style that detracts from its operational usefulness.

Psychographics (personality and life-style) provide a richer view on the market based on a more lifelike portrait of the consumer. Segments are readily identifiable, measurable and substantial. The accessibility and actionability of psychographic segments are their major benefits. Therefore, one of the most extensive uses of life-style segmentation is in the creation of advertisement messages

(Plummer 1974). Little research has been done on the validity of psychographics and the stability of psychographic segments. Lastovicka (1982) identified concepts that were relevant in assessing life-style trait validity, and Lastovicka (1982) and Lastovicka et al. (1990) provided comprehensive validation studies of life-style traits, which demonstrated that life-style typologies may show considerable validity. The responsiveness of segments based on general psychographic measures is weak, as was discussed above. General psychographic measures appear to be rather independent of any specific product-related behavior, but more likely show a relationship with extensive patterns of behavioral responses.

3.5 Unobservable product-specific bases

As the limitations of general psychographics for the explanation of product-specific behavior dawned, increasingly segmentation variables were developed that were specific to the consumer process being studied. The following groups of variables can be distinguished:
1. Product-specific psychographics,
2. Product benefit perceptions and importances,
3. Brand attitudes, preferences and behavioral intentions.

These groups form a hierarchy in the sense that the higher levels are influenced by all lower levels (Wilkie and Cohen 1977). The unobservable product-specific bases dominate the bases described in the previous sections, on actionability and responsiveness, in that they provide marketing management with clues on the filling in of marketing variables, and responsiveness.

3.5.1 *Product-specific psychographics*

Psychographic measures, assessing personality traits and life-style in relation to the choice behavior of the product under study, show a much stronger relationship with that choice behavior than do general psychographic measures. This is Wells' (1975) conclusion on the basis of a review of the literature. Correlation coefficients of 0.5-0.6 have been reported for these scales, whereas (Pearson) correlations of behavioral measures with general psychographic scales have been between -0.20 and 0.20.

Dhalla and Mahatoo (1976) identified three key areas of product-specific psychographics: value orientations, role perceptions and buying style.

Dickson (1982) claims that product-specific psychographic scales measure the different types of usage situations of the product that consumers with different life-styles are likely to encounter. Consequently, he recommends that the situational context should be assessed explicitly.

The product-specific psychographic measures thus add responsiveness to the conditions of effective segmentation met by general psychographics. Unfortunately, still little is known on the stability of the resulting segments.

3.5.2 ***Product benefit perceptions and importances***

3.5.2.1 *Perceptions*

Consumers' attitudes towards brands have been used as a basis for market segmentation since Yankelovich (1964). Three approaches are most frequently used to obtain the dimensions on which consumers perceive products: multidimensional scaling (Green and Carmone 1969), which develops perceptual maps from consumer judgments of the relative similarity of pairs of brands, and factor analysis (Hauser and Urban 1977) and discriminant analysis (Johnson 1971), which both estimate a limited number of perceptual dimensions from a large number of attribute scales rated by the consumers. Of the three methods for perceptual mapping, factor analysis is superior on predictive accuracy, interpretability and ease of use (Hauser and Koppelman 1979).

Frank, Massy and Wind (1972) summarized a number of studies that employed segmentation on the basis of perceived product attributes, and which showed that the resulting segments may be identifiable and substantial. The responsiveness criterion is also supported: Frank et al. (1972) also report a number of studies that demonstrate the relationship between attitudes and product usage (see also Assael 1970). In general, however, purchase behavior towards a product will be more strongly associated with a person's behavioral intention than with his perceptions of its attributes (Fishbein and Ajzen 1975). Little research has been done on the stability of perceptual segments, but cognitive states are bound to be much less stable than values (Wilkie and Cohen 1977).

In general, perceptual dimensions have low probity as segmentation bases, both from a theoretical and a strategic point of view (Dhalla and Mahatoo 1976, Wilkie and Cohen 1977, Howard 1985).

3.5.2.2 *Importances*

The concept of benefit segmentation was introduced into the marketing literature by Haley (1968). He argued that the different benefits people are seeking in consuming a product are the basic reasons for the existence of heterogeneity in choice behavior. More precisely, the importances of product benefits are the systematic sources of individual differences in buying behavior, and are thus the relevant bases for segmentation (Howard 1985). Before Haley, Yankelovich (1964) already proposed the identification of differences in consumer values - being the underlying causes of benefit importances (Howard 1985) - as the crucial issue in segmentation research. Similarly, Beldo (1966) segmented consumers on the basis of what he called 'their functional requirements of products', and Green, Wind and Jain (1972) suggested segmentation by benefit bundles.

In his review of segmentation, Wind (1978) listed benefits as a preferred segmentation basis for general understanding of a market, positioning, new product concepts, and advertising and distribution decisions, because of their actionability. Benefit segments are potentially identifiable and substantial as was demonstrated in a number of studies (Haley 1968, 1984, Beldo 1966, Calantone and Sawyer 1978). The responsiveness of benefit segments was investigated by Wilkie (1970). He demonstrated fair differences between benefit segments in brand purchases, and strong differences in attitudes and buying intentions (the study used self-rated benefit importances). Wilkie (1970) and Calantone and Sawyer (1978) showed that benefit segments can be made accessible, by relating them to demographic, socioeconomic and psychological characteristics and to usage occasions.

In their study on the stability of benefit segments, Calantone and Sawyer (1978) demonstrated that benefit segments are strongly consistent across independent (split-half) samples within a given time period. The segments' benefit importances remained intact over a period of two years. However, segment size, demographics and segment memberships changed in this period. As an explanation, the authors suggested that benefit importances may be situation-specific, while the usage situations consumers encounter change in time. The former hypothesis was supported empirically by Miller and Ginter (1979), which underlines the importance of the explicit consideration of usage context in benefit segmentation (Wilkie and Cohen 1977, Stout et al. 1977, Dickson 1982). The self-rated importances used by Calantone and Sawyer (1978) were put forward as an alternative explanation of the instability of segment membership, and less ambiguous assessment of benefit importance, such as by statistical estimation, was advised.

3.5.3 *Preferences and intentions*

Frank, Massy and Wind (1972) report that segmenting consumers on the basis of their intentions to buy is a fairly common practice (e.g. Sewall (1978), used this approach to segmentation). Buying intentions are, from a theoretical point of view, the strongest correlates of buying behavior discussed so far. Brand attitudes and preferences are expected to correlate somewhat less strongly with buying behavior (Wilkie and Cohen 1977). Frank, Massy and Wind (1972) report that preferences and intentions may result in identifiable, substantial segments. Accessibility has been demonstrated by relating intentions to general consumer descriptors, responsiveness by relating them to purchase behavior (Pieters (1989), reports a large number of studies that have investigated the relationship of attitudes and intentions with purchase behavior).

Little is known of the stability of segments based on measures of purchase predisposition. From a theoretical point of view, these segments are expected to show a greater temporal stability than segments based on the purchase behavior itself (Wilkie and Cohen 1977). The latter are less stable due to the scrambling effects of the purchase environment. In comparison with psychographics, benefit importances and benefit perceptions, preferences and intentions are less appealing from the point of view of actionability.

3.6 Conclusions

To evaluate the segmentation bases, the main criteria for effective segmentation are stability, accessibility, responsiveness, and actionability. (The bases discussed all have the potential of generating identifiable, and substantial segments.)

General consumer descriptors have generally shown a lack of success as segmentation bases in terms of responsiveness. Psychographic measures, especially when tailored to the product (class) in question, have exhibited a greater effectiveness with respect to these criteria, and dominate on accessibility and actionability. Situational variables have been shown to have great potential for segmentation. They should however be used in addition to, rather than instead of, consumer segmentation. In general, segments based upon purchase behavior or response to marketing variables may be less stable than are segments based upon behavioral intentions, although some measures of purchase behavior such as loyalty or usage frequency are quite stable. The resulting segments are actionable in the sense that they may indicate the magnitude of price changes or the number of promotional insertions in media, although they provide no indications for the contents of advertising messages, or the attributes of products to be modified. These bases are clearly superior in responsiveness. Segments based upon purchase intentions lack actionability, but satisfy the responsiveness criterion (although evidence on the association with response to marketing variables is lacking). The variables that are specific in the consumer decision process - benefit perceptions and importances - represent particularly actionable segment bases, providing diagnostics for marketing strategy. Benefit importances have, moreover, been shown to satisfy the stability, accessibility, and responsiveness criteria.

Segmentation is essentially a grouping task, for which a large variety of methods are available. The methods applied to segmentation research can be partitioned in two ways. The first way is to classify them into a priori and post hoc approaches (Green 1977, Wind 1978). A segmentation approach is called a priori when the type and number of segments are determined completely by the researcher, independent of the data collected. The researcher chooses the segmentation base and its cluster defining levels in advance. The objectively measurable segmentation bases (both general and product-specific) lend themselves best for this approach, which is frequently used in the microeconomic stream of segmentation research. If the complexities of the market can not be captured by the relatively small number of variables that can be accommodated in a priori segmentation schemes, a post hoc approach is more appropriate. An approach to segmentation is called post hoc when the type and number of segments are determined by the researcher on the basis of the results of the analysis of the data. Generally, the abundance of information provided by the large number of variables necessitates the use of multivariate statistical techniques to create a system from which conclusions can be drawn. This approach is more associated with the behavioral stream of segmentation research.

The second way to classify segmentation approaches is according to the type of statistical methods used. These methods can be classified into descriptive or predictive, according to whether they identify segments, or test the relationship of segmentation variables with purchase behavior or preferences or intentions (Sheth 1974). The predictive (functional, dependence) methods analyze the association between two sets of variables, where one set consists of dependent variables such as measures of purchase behavior or predisposition to purchase. The descriptive (structural, interdependence) methods analyze the mutual association across a set of segmentation variables, with no distinction between dependent or independent variables. The microeconomic segmentation stream uses mainly predictive methods, whereas the behavioral stream uses both types.

A classification of the statistical methods that are available for segmentation, according to these two criteria, results in 4 classes (Figure 4.1).

The choice of the method often critically determines the segments obtained. The classification of segmentation procedures (Figure 4.1), provides a framework for the evaluation of alternative procedures in relation to the marketing problem in question and the structure of the data. In the following sections the statistical methods are described that are available for segmentation. (Chapters 5 and 6 are concerned with specific approaches developed within the microeconomic and behavioral science streams.)

Figure 4.1. Classification of segmentation methods[1]

	A-priori	Post-hoc
descriptive	Contingency tables, log-linear models. (4.1.1)	Clustering methods: non-overlapping, overlapping and fuzzy. (4.2.1)
predictive	Cross tabulation, regression, logit and probit models, and discriminant analysis. (4.1.2)	Automatic interaction detector, Clusterwise regression. (4.2.2)

[1] The numbers in parentheses refer to sections containing a treatment of the methods in question.

A number of methods have been mentioned in connection to segmentation, but were not specifically designed to identify or test some grouping of consumers. These methods will not be considered here[1)]. (See e.g. Dillon and Goldstein (1984), or Marcia, Kent and Bibby (1988) for textbooks on multivariate methods. A recent review of multivariate analysis is provided by Schervisch (1987).)

4.1 A-priori approaches to segmentation

In a-priori segmentation methodology the type of segments and their number are determined by the researcher, while respondents are classified into segments on the basis of the data collected. With descriptive methods, associations between groupings from alternative bases may be assessed. Predictive methods are used to investigate if the bases are related to measures of purchase behavior, preferences or intentions.

4.1.1 *A-priori descriptive methods*

A simple way of displaying the associations between different segmentation bases is by cross-classifying the (categorical) segmentation variables in contingency tables. The representation of data in this way is simple, and associations can be tested. However, multidimensional contingency tables of a higher order are both conceptually and statistically more difficult to deal with.

Log-linear models were designed to study the interrelations between categorical variables that form a multi-way contingency table. The log-linear model assumes the cellfrequencies to follow a Poisson or Multinominal distribution, and models their expectations as a log-linear function of main effects and interactions of the classifying variables.

Green and Carmone (1977) suggested the use of log-linear models for segment congruence analysis. Purposes of the analysis are to test whether segments arising from alternative bases or methods exhibit mutual association (without depending on linearity assumptions), and to predict segment membership derived from an a-priori distinguished base on the basis of segment membership derived from other bases.

4.1.2 *A-priori predictive methods*

The predictive approaches require the prior specification of dependent and independent variables, and a model or class of models to describe the relation between the two. Two types, the forward and the backward approach can be distinguished (Wilkie and Cohen 1977). In forward approaches the a-priori defined segmentation variables (mostly general bases) are the independent variables, and it is investigated whether segments are related to the dependent variable, which is a measure of purchase behavior or of purchase predisposition. In the backward approach the segments are defined on the basis of purchase-related variables such as purchase volume or brand loyalty, and it is investigated whether differences in consumer characteristics exist between segments.

A large number of statistical methods are available that can be and have been applied for predictive segmentation. It is not the purpose here to describe all techniques in detail, but a short review will be provided (see also Wind 1978).

Cross tabulation

Bass, Tigert and Londsdale (1968) strongly advocated the use of this technique in segmentation studies. It entails tabulation of the dependent variable by the segmentation bases. The advantages are that nonlinearities and interactions can be isolated, and that simple t-tests can be used for significance testing. A difficulty, however, is the extension of the technique beyond two variables.

Regression [2)]

Regression permits multiple segmentation variables to be handled and their significance levels and partial contributions to be estimated. Effects are assumed to be linear and additive. The general regression model for P independent variables is:

$$y_i = \beta_0 + \Sigma_p \beta_p x_{ip} + \epsilon_i, \tag{4.1}$$

where y_i denotes the observation of the dependent variable for the i-th subject, x_{ip} represents the value of the p-th segmentation variable, β_0 and β_p ($p = 1 \ldots P$) represent the regression model parameters, and ϵ_i represents error terms that are assumed to be independent and normally distributed.

Bass et al. (1968) argued against the use of regression analysis in segmentation. They argued that regression has the individual and not the group (segment) as the unit of analysis, which is the case in e.g. cross-tabulation. Initially, this standpoint was confirmed by Morrison's (1973) analysis based upon a Negative Binomial model of purchase behavior, but Wildt (1976) and Beckwith and Sasieni produced (1976) counterevidence against Morrison's argument. They showed that in correctly specified regression models half of the variance may result from random variation in purchases.

Wildt and McCann (1980) proposed a regression model for market segmentation studies, which recognizes that consumption behavior on any given occasion fluctuates around some mean consumption level, according to a Poisson process. Variations in the mean consumption level across the population are assumed to be related to consumer characteristics. An iterative least-squares procedure was suggested for estimation. In each step, the responses are weighted with the covariance matrix of the residuals from the previous step, until convergence. It was shown that in situations involving a large number of purchases per subject, ordinary least-squares estimation of mean household purchase rate is efficient in comparison with the proposed procedure.

The multinomial logit and probit

Both logit and probit models can be used to relate segmentation variables to discrete dependent variables (such as number of purchases or brand loyalty). The latter are assumed to follow a multinomial distribution. Because of computational feasibility, the multinomial logit is the one most frequently used, despite of its drawbacks in application related to the independence of irrelevant alternatives (IIA) assumption (This assumption states that the ratio of choice probabilities between any two alternatives, which is based upon their utilities, is independent of the utility of any other alternative). Individual choices are assumed to derive from maximization of the (random) utility of subject i with respect to product j ($j = 1...J$). The fixed part of the utility, u_{ij}, is a linear function of the P segmentation variables, x_{ijp}:

$$u_{ij} = \Sigma_p \ \beta_p \ x_{ijp} , \qquad (4.2)$$

(β_1 denotes the intercept and $\mathbf{x}_1$ the unit vector). The assumption of a double exponential distribution of the error term ϵ_{ij} results in the logit model for the probability that individual i chooses product j, p_{ij} (in fact the IIA property arises from the assumption of independence of the errors):

$$p_{ij} = \exp(u_{ij}) / \Sigma_j \exp(u_{ij}) \tag{4.3}$$

Applications of the multinomial logit include those by Currim (1981), who estimated separate models relating consumer perceptions to discrete choice variables, within a number of situational and benefit segments. Gensch (1985) provided a likelihood ratio test statistic to test for segments. An extensive treatment of the multinomial logit model for specifying and testing segmentation models is provided by Ben-Akiva and Lerman (1985).

The multinomial probit (Daganzo 1979) results from the assumption that the random component of the utility is normally distributed, and the corresponding probability that consumer i chooses product j is:

$$p_{ij} = \int_{-\infty}^{u_{ij}} \Phi(z)dz \tag{4.4}$$

where $\Phi(z)$ is the probability density function of a standard normal variate. The multinomial probit allows for nonzero covariance between the random utilities and thus alleviates the IIA problem.
Rao and Winter (1978) applied the probit model in the context of segmentation.

Discriminant analysis

Discriminant analysis identifies linear combinations of the independent variables that best discriminate a number of segments (e.g. users/nonusers). The discriminant functions are derived from the eigenvalue decomposition of $\mathbf{W}^{-1}\mathbf{B}$, where $\mathbf{W}$ and $\mathbf{B}$ are the matrices of sums of squares and products within and between the segments. Discriminant rules are developed to assign consumers to segments. The method assumes multivariate normality of the predictors and equal covariance matrices within segments. Significance tests are available and stepwise procedures can be used to select the independent variables. Discriminant analysis is, in general, a technique for the description of segments rather than a method to predict which variables are useful for segmenting a market (see, e.g., Frank, Massy and Wind 1972). It is often applied in backwards approaches to segmentation.

4.2 Post-hoc approaches to segmentation

In the post-hoc approach, the identification of segments is dependent on the data collected and the methods applied. Descriptive post-hoc approaches entail the clustering methods. Predictive post-hoc (predictive pattern) methods combine predictive ability with a post-hoc grouping of respondents, and entail the automatic interaction detector and clusterwise regression. Clusterwise regression, being the major topic of this thesis, will be treated separately in Part 2.

4.2.1 *Post-hoc descriptive methods*

Cluster analysis is not one single set of cohesive techniques, but a collection of procedures each having its own advantages and disadvantages (Frank and Green 1968). Punj and Stewart (1983) elaborately reviewed the application of cluster analysis in marketing research. General reviews of cluster analysis were provided by Cormack (1971), Blashfield and Aldenderfer (1978), and Gordon (1987). A major distinction in methods can be made according to the type of partitioning obtained: nonoverlapping, overlapping or fuzzy (Hruschka 1986). In nonoverlapping clustering consumers belong to one and only one segment, in overlapping clustering consumers may belong to more than one segment and in fuzzy clustering consumers may any possess a degree of membership (or a probability of membership) in different segments.

Two major types of nonoverlapping cluster techniques can be distinguished: the hierarchical (agglomerative) and the nonhierarchical (or partitioning) methods. According to Baker (1988), the use of agglomerative methods corresponds with the stance of the behavioral school. Individuals are regarded as potentials markets which are to be combined into groups. The economists' undifferentiated demand schedule would be the logical starting point for a partitioning approach.

4.2.1.1 *Nonoverlapping clustering*

Hierarchical clustering

Hierarchical clustering methods start with single-subject clusters, and link clusters in successive stages of the algorithms on the basis of similarities between the subjects in the clusters. A variety of measures can be used to characterize the similarity between objects.

For metric data, distance measures are used, for nonmetric data matching measures (e.g., Mardia et al. 1988).

On the basis of the way the similarities between subjects are used to link clusters, different hierarchical methods can be distinguished:

1. *Single linkage.* Clusters are joined on the basis of the maximum similarity with one of the cluster members.
2. *Complete linkage.* Clusters are joined on the basis of similarity with all current cluster members.
3. *Average linkage.* Clusters are joined on the basis of average similarity with the current cluster members (alternatively, average similarities between two or more of the closest cluster members can be used). Weighted *average*, *median* or *centroid*-based similarities have been used. The latter two have been shown to have undesirable properties.
4. *Minimum variance linkage.* Clusters are joined in such a way that the trace of W is minimized. W is the pooled within-cluster sum-of-squares matrix (Wards' method).

In addition to the above methods, a number of approaches have been presented that cluster objects and derive respective variable weightings indicating the importances of the variables in clustering. De Soete et al. (1985) summarize the literature on this topic and present a method that simultaneously clusters objects and derives the weightings of the variables, using an alternating least-squares algorithm.

Nonhierarchical clustering

The nonhierarchical, or partitioning, methods start from an initial division of the subjects into a predetermined number of clusters and reassign subjects to clusters until a decision rule terminates the process. The methods differ with respect to:

- the starting partition (random or based on a-priori information; alternatively, several random partitions may be used),
- the updating of cluster centroids (after each membership move or after a complete pass through the data),
- the criterion to be optimized (trace of **W**, trace of $\mathbf{W}^{-1}\mathbf{B}$, determinant of **W**, largest eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$. (**B** is the between-cluster and **W** the pooled within-cluster covariance matrix),
- the type of reassignment process.

Minimizing the determinant of **W** can be shown to yield Maximum Likelihood estimates of the grouping of consumers into segments, if the observations are assumed to follow a multivariate normal distribution (Scott and Symons 1971).

The frequently used method of K-means clustering is equivalent to minimizing the trace of W. DeSarbo et al. (1984) extended the K-means algorithm to simultaneously render weight of the variables indicating their relative importance to clustering, and to allow for the analysis of several different groups of a-priori weighted variables. The usefulness of their nonhierarchical SYNCLUS algorithm to market segmentation lies in the screening of candidate segmentation variables and in exploring the robustness of cluster structure to the importance of alternative sets of variables (such as psychographics and demographics).

Evaluation of hierarchical and nonhierarchical methods

Based on a review of the literature, Punj and Stewart's (1983) main conclusions with respect to the comparison of the various cluster analysis techniques are as follows:[3)]

- Nonhierarchical methods are superior to hierarchical methods, especially if nonrandom starting partitions are used. They are more robust with regard to outliers, the distance metric chosen, and the presence of irrelevant attributes.
- Among the nonhierarchical methods, minimization of det(**W**) is superior.
- Hierarchical methods produce better results when Pearson product moment or intra-class correlation coefficients are used as similarity measures, or when the

data are standardized prior to clustering. This reduces the sensitivity to the presence of outliers.

- The performance of the hierarchical procedures tends to deteriorate at higher levels of coverage, when more observations are assigned to clusters.
- The performance of the hierarchical procedures tends to deteriorate when irrelevant variables are included in the analysis (see also Funkhouser 1983).
- Among the hierarchical procedures, Wards method outperforms the other ones while average linkage rates second.

4.2.1.2 *Overlapping and fuzzy clustering*

A major problem concerned with the clustering procedures applied in segmentation research is the assumption of nonoverlapping segments. The assumption implies that subjects within a segment do not share properties with subjects in other segments. The validity of an approach in which consumers can belong to only one segment has been questioned by Arabie (1977), Shepard and Arabie (1979) and Arabie et al. (1981). There is no reason for consumers to occur in groups or segments, but the limitations in processing a multivariate continuum impose nonoverlapping classification as a conceptually attractive abstraction of reality (Johnson, 1971).

Beldo (1972) already pointed out that intraindividual segmentation may be appropriate as consumers may want different benefits from the same product, depending, for example on the usage situation. A similar argument was set forth by Dickson (1982). Methodological constraints have long precluded the consideration of the possibility of consumer membership in more than one segment, and have restricted conventional segmentation to less realistic but technically simpler identification of nonoverlapping segments (Arabie et al. 1981).

Overlapping clustering

In overlapping clustering objects can either be or not be a member of each of the clusters identified. Reviews of overlapping clustering were given by Cormack (1971), and Arabie (1977).

Shepard and Arabie (1979) suggested the overlapping clustering model ADCLUS for the analysis of a symmetric two-way matrix containing pairwise similarities. Arabie and Carroll (1980) devised an algorithm for fitting the model (MAPCLUS); an application to market research was given by Arabie et al. (1981). Carroll and Arabie (1983) generalized the ADCLUS model. The INDCLUS model proposed by them deals with the analysis of different similarity matrices for a number of subjects.

The GENNCLUS algorithm of DeSarbo (1982) deals with nonsymmetric similarity data, and allows for (overlapping) clustering of both modes (e.g., two different sets of stimuli) of this matrix.

CONCLUS is a clustering methodology developed by DeSarbo and Mahajan (1984), which was devised to perform constrained classification. It operates on a matrix of derived distances between objects. In CONCLUS, constraints can be imposed on the allowable classifications and their characteristics. The constraints that can be imposed include nonoverlapping (both hierarchical and nonhierarchical), overlapping and fuzzy clustering. Constraints on cluster membership may force two or more objects within the same or into different clusters. Restrictions on the size of clusters, and differences in the size or deviations between clusters may be imposed.

The ADCLUS, INDCLUS, GENNCLUS and CONCLUS models provide useful approaches to portraying a discrete structure of similarities. To yield overlapping consumer segments, the methods must be applied to derived similarities between subjects. The computational requirements for some of the iterative algorithms discussed above may be excessive for real-life segmentation problems with larger sample sizes (Arabie et al. (1981) recommended that no more than 30 observations should be used.)

Fuzzy clustering

In fuzzy clustering, two distinct approaches can be distinguished. The first of these is based upon the assumption that the data arise from a mixture of distributions (McLachlan and Basford 1988). Two assumptions are most frequently used, that of the normal and of the Bernoulli distribution. The method corresponding to the latter assumption is known as Latent Class Analysis. Due to the distributional assumptions made, the mixture methods allow for maximum likelihood estimates of their parameters. Objects are classified into clusters on the basis of a-posteriori calculated probabilities.

The second approach to fuzzy clustering is based upon the theory of fuzzy sets (Zadeh 1965). Fuzzy set theory assigns gradual membership to objects indicating their nearness to a class (Hruschka 1986). In various forms of fuzzy clustering algorithms, memberships are estimated by minimizing objective distance criteria.

The mixture approach: Latent Class Analysis (LCA)

LCA is a statistical method for the analysis of multi-way contingency tables (Goodman 1974). It attempts to explain the observed associations between the variables that make up the table by unobservable underlying factors. When these unobservable factors are introduced into the table, the original classifying factors of the table are independent at each level of the latent factor. LCA assumes that the association between two or more categorical variables in a table is due to a mixing of unobservable clusters and that within clusters the variables are independent. The purpose of LCA is to estimate the latent classes (clusters) and the probabilities of each individual to be a member of the clusters. In the latent class model, the observed proportion in the l_k-th cell of the table, defined by the l-th category of the k-th classifying variable (k = 1...K, denotes the number of variables and L_k the number of classes of variable k), is approximated as:

$$\hat{P}_{l_1 \ldots l_k \ldots l_K} = \Sigma_i \, \pi_i \, \pi^i_{l_1} \ldots \pi^i_{l_k} \ldots \pi^i_{l_K}. \qquad (4.5)$$

where π_i is the probability that an individual is a member of a latent class i (i = 1...c), and $\pi^i_{l_k}$ is the conditional probability of being in the l-th category of the k-th observed variable given latent class i. The latent class parameters are estimated by maximizing likelihood.

Green et al. (1976) suggested the use of LCA for segmentation and Grover and Srinivasan (1987, 1989) applied it to simultaneous segmentation and market structuring. These approaches will be described in more detail in Chapter 5.

The mixture approach: normal distributions

This approach to fuzzy clustering starts from the assumption that the sample in question arises from a mixture of normal densities of underlying clusters (Wolfe 1970, McLachlan and Basford 1988). The number of clusters, c, is assumed to be known. The sample $x_1,...,x_n$ of measurement vectors for subjects j ($j=1...n$), is assumed to be drawn from the normal densities of the underlying groups, mixed in unknown proportions $\pi_1,..., \pi_c$, ($\Sigma_i \pi_i = 1$; $i=1...c$):

$$x_j \sim \Sigma_i \pi_i f_i(x) \tag{4.6}$$

where $f_i(x)$ refers to the multivariate normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{S}_i$.

The method of maximum likelihood is used to estimate the parameters of the underlying distributions. After these parameters have been estimated, the posterior probability that individual j belongs to cluster i can be calculated using Bayes' rule:

$$\text{prob}(j \in i) = \hat{\pi}_i \hat{f}_i(y_j)/\Sigma_h \hat{\pi}_h \hat{f}_h(y_j) \tag{4.7}$$

where $\hat{f}_i(y_j)$ is the normal density function, evaluated at y_j and $h=1...c$. Whereas this procedure is applicable only to two-way data, Basford and McLachlan (1985) extended it to three-way data sets, in which one mode is to be clustered on the basis of the other two simultaneously. DeSarbo and Cron (1988) developed a conditional mixture procedure, in which the expectations of the mixtures are postulated to be linear functions of explanatory variables. (This method, being a clusterwise regression procedure, will be described in more detail in Part 2.)

The number of clusters has to be assessed empirically in applications of these mixture methods, and statistics have been suggested to guide its determination (McLachlan and Basford 1988).

Fuzzy c means and varieties (FCM, FCV)

The concepts of fuzzy sets and partial set membership were introduced by Zadeh (1965) as a way to handle imprecision in mathematical modeling. Ruspini (1969), Dunn (1974) and Bezdek (1974) recognized the applicability of Zadeh's concepts to clustering problems. Dunn and Bezdek proposed the fuzzy c means (FCM) algorithm, which partitions data into clusters by iteratively minimizing the criterion:

$$J_{Mm} = \Sigma_j \Sigma_i u_{ij}{}^m D_{ij}, \tag{4.8}$$

under the constraints $\Sigma_i u_{ij} = 1$ ($i = 1 \ldots c$, $j = 1 \ldots n$). The exponent m is a fuzzy weight parameter ($m > 1$) that influences the extent to which clusters are fuzzy (as m approaches 1, a hard partition is obtained), $\hat{u}_{ij}$ denotes the membership of subject j in cluster i ($0 < u_{ij} < 1$) and $\hat{D}_{ij}$ denotes the (Euclidian) distances of observation vector j, x_j, from the centroid of cluster i ($\hat{x}_i$):

$$\hat{D}_{ij} = (x_j - \hat{x}_i)'(x_j - \hat{x}_i). \tag{4.9}$$

The purpose of the FCM algorithm is to estimate the cluster centroids and memberships. In the iterative algorithm, resulting from the minimization of J_{Mm}, the $\hat{x}_i$ and $\hat{u}_{ij}$ are calculated from:

$$\hat{x}_i = \Sigma_j \hat{u}_{ij}{}^m x_j / \Sigma_j \hat{u}_{ij}{}^m, \tag{4.10}$$

$$\hat{u}_{ij} = 1/\Sigma_h (\hat{D}_{ij}/\hat{D}_{hj})^{1/(m-1)} \tag{4.11}$$

A generalization of the fuzzy c means clustering algorithms was developed by Bezdek et al. (1981 a,b), fuzzy c varieties (FCV). FCV identifies not only round clusters, but also clusters with any shape that can be described as a linear surface. (In FCV, the above formulas for the cluster centroids $\hat{x}_i$ and memberships $\hat{u}_{ij}$ apply, but now the $\hat{D}_{ij}$ equal the eigenvectors of the maximum eigenvalue of the within cluster fuzzy scatter matrix.)

Hruschka (1986) reviewed a number of different fuzzy clustering algorithms, including methods that use a (dis)similarity matrix between subjects, methods that are based upon the concepts of affinity and neighborhood, and nonmetric and hybrid methods. An empirical comparison of a number of these methods, and a nonoverlapping and an overlapping (ADCLUS) procedure for market segmentation demonstrated the superiority of the fuzzy methods in terms of internal validity (percentage of variance accounted for) and external validity (relationship of memberships with variables not used in the clustering).

4.2.2 *Post-hoc predictive methods*

The post-hoc predictive methods combine grouping of consumers with predictive ability. Dependent variables are mostly measures of purchase behavior or purchase predisposition. These methods, which have been called predictive pattern techniques (MacLachlan and Johansson 1981), provide a powerful approach to segmentation. Clusterwise regression, also a predictive pattern method, will be treated in Part 2. Will focus here on the other major technique within this class, the automatic interaction detector.

Automatic interaction detector (AID)

AID is a method for identifying interactive effects of independent (segmentation) variables on a dependent variable. AID divides the sample into groups on the basis of the independent variables. These variables have to be categorized. The algorithm considers each variable in turn, and uses all possible dichotomizations. It is decided which split for which variable results in the largest reduction in the residual sum of squares in the regression of the dependent variable. The two groups formed are each a candidate for further splitting in the next step, in which all but the variable already entered are considered in turn. The process continues until user-imposed constraints are met, or predictive accuracy can not be improved (Assael, 1970).

AID was generalized to deal with multiple dependent variables (MAID, MacLachlan and Johansson 1981), with nominal dependent variables (THAID, see Wind 1978), and with categorical dependent variables (CHAID, Kass 1980).

A related binary tree method is classification and regression trees (CART, Breiman et al. 1984), which is nonparametric.

The advantages of AID are the simultaneous grouping and prediction it provides, as well as the ease of interpretation and communication to managers. A number of serious problems concerning the technique have been identified however (Doyle and Fenwick, 1975). Sample sizes have to be large, 1000 and over. If predictors are correlated, the order of appearance is hardly an indication of relative importance, and the resulting trees are instable. Stability of the trees is also affected by noise in the data (although end groups are reported to be stable). Significance tests are not provided for. Although AID claims to detect non-linearities and interactions, only interactions of two variables that show a main effect will be detected. Cross-validation of solutions is essential, as Doyle and Fenwick (1975) showed that the predictive value of the model and its robustness in cross-validation is often not satisfactory. They advice, therefore, not to trust results which are not validated and to use the technique as an exploratory method before applying regression.

4.3 Conclusions

The classification of segmentation procedures developed provides a framework for the evaluation of alternative segmentation procedures in relation to the requirements of marketing management and the structure of the data collected.

A-priori segmentation methods are appealing because of the simplicity of their concept and the ease of interpretation of the results. The a-priori methods have been predominantly used in the microeconomically oriented stream of segmentation research. This stream traditionally focused upon 'hard' data, such as the observable segmentation bases, that lend themselves best to a-priori segmentation. Partially, however, the a-priori approach to segmentation within this stream of research resulted from a lack of statistical procedures to identify segments post hoc. Only recently, developments in this area were made, as will be described in the next chapter.

The post-hoc methods are appealing because they deal with the complexity of the market situation, without the restriction of a-priori convictions about segments and their boundaries. The post-hoc approach has predominantly been associated with the behavioral stream of segmentation research. The advent of multivariate statistical methods and the availability of computer programs encouraged the application of a large variety of statistical methods to segmentation problems, perhaps not always with an adequate commitment to the assumptions underlying the methods. The use of these methods should be based on careful consideration of their representation of the structure of the data in relation to the assumptions involved.

Segmentation is inherently a grouping task, to which methods devised for grouping and classification should be applied. From the available nonoverlapping clustering methods, the partitioning methods have been shown to be superior. The use of nonoverlapping clustering seems to be unnecessarily restrictive in many cases. Subsets of subjects may share unobserved properties to any degree, which will result in partial cluster membership and overlapping or fuzzy clusters. The available overlapping clustering methods have been developed for the analysis of similarity data and have been applied to market structuring rather than to segmentation. Fuzzy clustering methods are particularly interesting, in that they allow for within-person segmentation, and may thus account for subjects belonging to a number of segments, for example in relation to different usage situations.

Within the class of post-hoc approaches the predictive pattern techniques have a great potential for segmentation research. These methods combine the advantages of post-hoc identification of segments with prediction of the dependent measure of interest, such as preference or purchase intention. Within this class of methods AID has been used relatively often, although it has some serious shortcomings.

5 APPROACHES TO SEGMENTATION WITHIN THE MICRO-ECONOMIC SCHOOL

In the microeconomic stream mathematical models are used to describe and explain consumers' purchase behavior. Leeflang (1974) classified the models used according to their purpose:

1. Descriptive models explain consumer purchase behavior without explicitly taking the effects of marketing variables into account. These models predict the (cross-sectional) behavioral level.[1)]
2. Demand response models assess the reactions of consumers to the marketing variables used by firms. The models are concerned with behavioral response in time.
3. Policy models. This type of models entail a profit function which is maximized to yield the optimal allocation of marketing instruments. Policy models are based upon demand response models.

Segmentation approaches within the microeconomic stream can be classified according to the type of models that are used. The above classification of models will be used for this purpose.

5.1 Descriptive approaches

A large variety of descriptive approaches have been used in the segmentation literature. We will briefly mention a number of these.

Lessig and Tollefson (1971) proposed a two-stage clustering approach, in which consumers were grouped first according to measures of their buying behavior and then, within the resulting segments, on the basis of general characteristics. An empirical application failed to support the validity of the method.

Sexton (1974) proposed an approach in which consumers were clustered (hierarchically) on the basis of purchase habit variables, as these were expected to relate to sensitivity to marketing policies. The segments identified exhibited significant differences in response to market variables.

Frank (1972) and Frank and Strain (1972) used canonical correlation to relate purchase data to segmentation variables (demographic, socioeconomic and psychographic). Segments were obtained by cross-classification of the scores of the segmentation bases on the canonical variates. The segments obtained showed considerable differences in attitudes and choice behavior. This approach provides a step in the direction of simultaneous grouping of consumers and prediction of intentions or behavior. A disadvantage of the approach is that the grouping task involved the rather ad-hoc procedure of identification of segment boundaries by visual inspection.

Assael (1970) and Assael and Roscoe (1976) used AID to identify consumer groups by demographic, socioeconomic and attitudinal variables. The resulting segments showed considerable differences in purchase behavior. The stability of the resulting segments was not investigated; the limitations of AID in this respect have been discussed above.

Backward approaches to segmentation involve the a-priori identification of segments into which consumers are grouped on the basis of purchase behavior itself; mostly, brand loyalty is used as a segmentation basis.

Starr and Rubinson (1978) classified consumers into brand loyalty segments on the basis of the choice probabilities for their primary brand. Blattberg and Sen (1974, 1976) also defined brand loyalty segments a priori. Markov models (see e.g. Massy, Montgomery and Morisson 1970) were developed for each segment to explain brand loyalty. Each consumer was classified into the segment the model of which had the highest probability of generating his purchase history (using a Bayesian procedure to calculate these posterior probabilities). The a-priori identification of the brand loyalty segments according to the procedures outlined is quite cumbersome, however, involving the visual inspection of each consumer's purchase history. Blattberg et al. (1978) assigned consumers to loyalty groups and developed a model to explain the behavior of the deal-prone segment.

Green et al. (1976) suggested the application of latent class analysis to brand switching data, to obtain post-hoc segments with different brand switching behavior. Grover and Srinivasan (1987) developed this further into an approach for simultaneous segmentation and market structuring. Their method is based on the cross-classification matrix of brands chosen on two purchase occasions.

Heterogeneity of brand choice probabilities across consumers is assumed to be captured adequately by a brand-loyal segment for each brand i, containing a proportion V_i of the consumers, and K brand-switching segments, containing a proportion W_k (k=1...K) of the consumers. The probabilities of choosing brand i for switching segment k (p_{ik}) are assumed to be constant over the time period of the study. The theoretical proportions of consumers in the market buying brand i on one occasion and brand j on another (S_{ij}), and the proportion of consumers buying brand i on both occasions (S_{ii}) were derived:

$$S_{ij} = \Sigma_k W_k p_{ik} p_{jk} \tag{5.1}$$

$$S_{ii} = V_i + \Sigma_k W_k p_{ik}^2 \tag{5.2}$$

These equations were shown to correspond to the latent class model equations, which can thus be used to estimate the proportions of consumers in the segments (V_i and W_k) as well as the within-segment probabilities of purchasing the brands (p_{ik}). The number of switching segments, K, has to be determined empirically. The method was generalized to a response type model that accounts for nonstationarity in the within-segment market shares over the time horizon considered (Grover and Srinivasan 1989). Considerable diagnostic advantages are connected with these approaches, as they provide insight into the patterns of competition of brands within segments. The methods proposed by Grover and Scrinivasan were the first post-hoc segmentation methods within the microeconomic school.

5.2 Demand response segmentation

The demand response model relates changes in the number of purchases of a brand (Wildt and McCann 1980), or the brand's market share (Frank and Massy 1965, Sexton 1974, McCann 1974), to changes in the marketing instruments employed by the firm and those of its competitors. Econometric techniques based on multiple regression are used to estimate the parameters of the models from pooled cross-sectional and time series data, within a-priori defined segments.

Whereas most of the models used have not explicitly considered competitive reactions across segments, Plat and Leeflang (1988) developed a model to deal with the reactions of competitors, operating on different segments. The model assumes a-priori defined segments. It offers insights into the competitive reactions with respect to the market instruments, on segmented markets.

Whereas the studies mentioned above have used a-priori segmentation designs with objectively measurable segmentation bases, a few studies have been reported that used demand response itself as a basis for (post-hoc) segmentation.

Assael and Roscoe (1976) used changes in demand as a dependent variable in AID analysis. The segments, identified from demographic descriptors, showed considerable differences in responsiveness. Elrod and Winer (1982) related the amounts purchased to relative price, using a separate response model for each consumer. A number of local measures of the response functions (elasticities, coefficients, etcetera) were used to group consumers into segments, using hierarchical clustering.

The potentially powerful approach to segmentation on the basis of demand response, has in practice been used frequently in combination with rather weak a-priori segmentation designs and general consumer descriptors as segmentation bases (Frank 1972, McCann 1974). The ideal approach of estimating demand response at the individual level, and grouping consumers with homogeneous response into segments can be one of the most profitable approaches to market segmentation. However, in practice, the grouping of consumers on the basis of their individual response to marketing variables is frequently not feasible. Frank, Massy and Wind (1972) report one substantive effort to deal with individual demand functions, the results of which "do not given one confidence in the efficacy of a fully disaggregative approach".

Recently, however, Kamakura and Russell (1989) proposed a probabilistic (latent class) choice model for market segmentation that alleviates the problem of estimating the response functions at the individual level. The approach is based upon a multinomial logit model of choice, with price as the independent variable. The existence of $i = 1 \ldots c$ homogeneous segments is assumed, with relative sizes π_i. The probability of choosing brand k is:

$$p_k = \Sigma_i \pi_i \exp(u_{ik} + \beta_i x_{jkt})/\Sigma_{k'} \exp(u_{ik'} + \beta_i x_{jk't}) \tag{5.3}$$

where x_{jkt} is the price of brand k (k = 1...K) for consumer j (j = 1...J) at time t, β_i is the price parameter in segment i, and u_{ik} is the intrinsic utility of brand k in segment i. The model thus assumes that the observed choices result from a mixture of underlying multinomial distributions. The posterior probabilities of membership in a particular segment are estimated using a Bayesian procedure:

$$\text{prob}(j \in i) = \pi_i L_{ij}/\Sigma_h L_{hj} \tag{5.4}$$

where L_{ij} is the likelihood of consumer j's purchase history given that he is in segment i. The competitive structure of the market is represented by the predicted within-segment brand choice shares. The model allows the identification of the effects of different price sensitivities within segments on the choice shares. The result is a description of market structure that links the pattern of brand switching to the magnitude of own- and cross-price elasticities within segments.

The Kamakura and Russell (1989) approach is a powerful approach to segmentation within the microeconomic school, and is the first approach to alleviate the problems inherent in grouping consumers in a two-stage procedure: the estimation of purchase response at the individual level and the local measure of demand to be chosen for the aggregation (e.g. Blozan and Prabhacker 1984). The procedure simultaneously yields the demand response functions within segments, as well as the (posterior) probabilities for each consumer that he belongs to these segments. A possible disadvantage of the method is related to the IIA assumption of the multinomial logit model.

5.3 Policy-oriented segmentation

A limited number of authors have followed this approach to segmentation. Most of these have had the objective of obtaining the optimal allocation of marketing variables to a-priori defined segments, from the maximization of a profit function.

Whereas most of the methods are based on demand response functions, Martin and Wright (1974) developed an AID-like algorithm, SIMS, from a simple cost-profit formulation. The analysis identifies profitable segments, but actually differentiating between some of these segments would be operationally difficult or costly. Besides, as the authors indicate, the cost function used is quite simple. The method is burdened with the same disadvantages as AID.

Winter (1979) proposed a two-stage approach to cost-benefit segmentation. At the first stage, segments are formed using a disaggregative (benefit) approach, while at the second stage the segments are aggregated on the basis of the profits resulting from a number of possible marketing mix strategies aimed at each segment (using an integer programming method). The method however requires the (demanding) estimation of the segment-marketing profit matrix, which in practice involves the specification of the mixes on discrete levels and a judgmental reduction of the number of mixes to be considered.

Other approaches within this school of segmentation research have typically been based upon demand response models (Dorfman and Steiner 1954).

The model for determining the allocation of price and promotion to market segments, developed by Claycamp and Massy (1968) and Frank, Massy and Wind (1972), starts from a profit function, Π, which is defined as total revenues, R, minus total production costs, C. Cost is a function of production, transportation and selling, at the quantity demanded within the segments. The demand within segment i ($i = 1 \ldots I$) is a function of price (p) and the number of promotional units (contained in the vector $\mathbf{x}_i$), inserted in the available media. The cost of promotion directed to each of the segments is incorporated in the cost function ($\mathbf{c}_i$ denotes the vector of the costs per unit of each of the promotional types for segment i). The profit function is formulated as follows:

$$\Pi = \Sigma_i \, p \, D_i(p, \mathbf{x}_i) - C(\Sigma_i \, D_i(p, \mathbf{x}_i)) - \Sigma_i \, \mathbf{c}_i' \mathbf{x}_i. \qquad (5.5)$$

Whereas in Eq. 5.5 the price to each consumer is the same, the formulation may be extended to include different prices for different consumers. The profit maximizing solution is obtained by setting the derivatives of Π equal to zero and solving for p and $\mathbf{x}_i$. Alternatively, the profit maximizing solution can be obtained for a fixed promotional budget, B, by minimizing Eq. 5.5 under the budget constraint.

The model was extended to include real-world constraints. It was assumed that media coverage is known only for groups of consumers - the microsegments -, and that the demand functions are known only for groups of microsegments with approximately the same response to promotion (macrosegments). The resulting expression for the extended model is similar to Eq. 5.5, the summations now being across macrosegments, microsegments within macrosegments and consumers within microsegments respectively. The approach, being concerned with the allocation of promotion to segments, starts from microsegments that are identified a priori, for example on the basis of a cross-classification of general consumer descriptors for which media exposure is known. For the grouping of microsegments into macrosegments, the authors suggest clustering of microsegments' response derivatives. Empirical applications of the approach are lacking. Nevertheless, the model has contributed (Lilien and Kotler 1983) in that it has shown segmentation to be an aggregative process in which variance of the within-group response function is minimized.

The model was extended by Mahajan and Jain (1978) to allow for simultaneous segment identification and optimal allocation of resources. Their model entails the maximization of a profit function under a number of possible constraints, including budget constraints and constraints on (nonoverlapping) segment membership, such as the maximum/minimum number of subjects allowed in a segment and maximum dissimilarity of subjects in a segment. The dynamic segmentation problem is to maximize profit, which is formulated as:

$$\Pi = \Sigma_i \Sigma_j b_{ij} p_{ij} D_j(p_{ij}, x_i) - C(\Sigma_i \Sigma_j b_{ij} D_j(p_{ij}, x_i)) - \Sigma_i c'_i x_i. \tag{5.6}$$

The notation is as used above, j denotes consumers ($j = 1...n$), b_{ij} indicates whether consumer j is in segment i, and p_{ij} is the price charged to consumer j if he is assigned to segment i. The allocation of promotional budget and the determination of price to be charged are obtained by minimizing Eq. 5.6 under the constraints. If the assignment of subjects to segments (contained in the b_{ij}) has been determined a priori, the model proposed by Frank, Massy and Wind (1972) becomes a special case of this model. The dynamic formulation of Mahajan and Jain allows, however, for the post-hoc determination of segments through the estimation of those b_{ij} that maximize Eq. 5.6. The model requires knowledge of every consumer's demand function, and estimation procedures were not provided. (The model was also extended to include real-world constraints, including the matching of media profiles with segment profiles.) The model awaits empirical applications.

Another procedure for the post-hoc estimation of segments on the basis of a dynamic profit formulation was provided by Tollefson and Lessig (1978). They proposed an aggregative approach, in which two segments are aggregated when the loss in profits from marketing to the merged segments as compared to marketing to the separate segments is minimum. The procedure is based upon the contribution of segment i to the firm's profit $\Pi_i(x_i)$. This contribution is a function of the firms marketing activity variable directed to the segment, x_i. The profit function is formulated as follows:

$$\Pi_i = p\, D_i(x_i) - C\,(x_i), \tag{5.7}$$

where the notation is as above. The profit reduction due to aggregation of segments h and i is:

$$\Delta_{hi} = \Pi_i(x_i^o) + \Pi_h(x_h^o) - \Pi_{hi}(x_{hi}^o), \tag{5.8}$$

where x_i^o is the value of the marketing variable that maximizes the profit from segment i, and $\Pi_{hi}(x_{hi}^o)$ is the profit associated with the allocation of marketing effort to the aggregated segments h and i. Starting with individual consumers as segments, segments are aggregated successively in such a way that at each stage Δ_{hi} is minimum. The authors claim that the formulations correspond directly to applied segmentation problems such as direct mail, but the procedure has only been applied to simulated data with a single marketing activity variable. In practice, the profit criterion is computationally infeasible, since optimal values of the decision variables and profits must be determined for each segment and each combination of two segments, at each stage of the procedure (Elrod and Winer 1982).

5.4 Conclusions

Descriptive approaches to segmentation have been quite heterogeneous. A number of authors (Frank 1972, Frank and Strain 1972, Assael 1970, Assael and Roscoe 1976) have applied simultaneous classification and prediction approaches, including canonical correlation and AID. Whereas most of the approaches using brand loyalty as a segmentation basis employ an a-priori identification of segments (Blattberg and Sen 1976, Starr and Rubinson 1978), Grover and Srinivasan (1987, 1989) developed a post-hoc approach, based on latent class analysis.

In approaches using demand response models, a-priori segmentation has traditionally been popular because of difficulties in estimating demand functions at the individual level, and thereby in grouping consumers on the basis of these estimates (Frank and Massy 1965, Frank 1972, McCann 1974). Again, a latent class model has provided a solution by simultaneous segmentation and estimation (Kamakura and Rusell 1989). The methods of Kamakura and Rusell (1989) and Grover and Srinivasan (1987, 1989) both provide insight into the patterns of competition between brands within segments. The method of Kamakura and Rusell, moreover, relates these patterns to different price sensitivities within segments. (Grover and Srinivasan, 1989, accomplish this in a two-stage procedure.)

Policy-oriented segmentation research has yielded relatively few procedures that have been implemented in segmentation practice. The models developed have served as a theoretical underpinning of segmentation research, and have guided the thought about normative ideals for segmentation. Whereas the a-priori segmentation designs used in some of the models can generally not be expected to yield segments that require vastly different allocation of marketing variables (Frank, Massy and Wind 1972), the application of post-hoc approaches has been hampered by difficulties in the estimation of individual demand and excessive computational requirements (Mahajan and Jain 1978, Tollefson and Lessig 1978).

APPROACHES TO SEGMENTATION WITHIN THE BEHAVIORAL SCHOOL

The behavioral stream of research tries to identify consumer segments on the basis of the differences in psychological processes underlying choice behavior. In the present chapter we will concentrate upon the segmentation approaches that are linked to the (compensatory) multiattribute models developed for describing consumers attitude structure[1]. These models have received considerable attention in the marketing literature because of their relevance from a managerial point of view (Wilkie and Pessemier 1973).

The expectancy value theory has been used in attitude research to identify the determinants of motivated behavior. This theory states that a tendency to act is the result of the strength of the expectancy that the act will be followed by a consequence, and the value of that consequence to the individual. Rosenberg (1960) and Fishbein (1967) have proposed models of attitude structure. The adequacy-importance model, is the most widely used model appearing in the consumer behavior literature (Bass and Talarzyk 1972, Cohen, Fishbein and Ahtola 1972). In this model (called the vector model) products or brands are postulated to be viewed by consumers as a bundle of attributes. A consumer's overall affect (preference, attitude intention) for a product equals the sum across attributes of that consumers' beliefs (perception) of the extent to which the attributes are offered by the brand, weighted with the importance of the attributes to the consumer[2].

In the ideal-point model (Carroll 1972), an offshoot of the vector model, each consumer is assumed to have an ideal level of each attribute at which the overall evaluation of the product is maximal, whereas it is assumed to decrease as the distance from the ideal point increases.

The assessment of both importance weights and brand perceptions falls into three categories: direct rating, compositional estimation, and decompositional estimation. For brand perceptions compositional methods for assessment entail factor analysis and discriminant analysis, whereas the major decompositional method is multidimensional scaling.

Importance weights are similarly assessed by either direct rating or compositional or decompositional estimation. In compositional estimation of importances overall evaluations are related to perceived dimensions (using vector or ideal-point models; e.g. PREFMAP, Caroll 1972; MONANOVA, Green and Wind 1973; LINMAP, Shocker and Srinivasan 1974; regression, e.g. Hauser and Urban 1977; multinomial logit or probit models, e.g. Rao and Winter 1978). With decompositional estimation overall evaluations are decomposed into the part-worths of attribute levels (using the conjoint model, Green and Srinivasan 1978).

6.1 The two-stage approach to benefit segmentation

Benefit importances have been postulated to be the source of systematic and stable differences in buying behavior among consumers (Wilkie and Cohen 1977, Howard 1985). In benefit segmentation consumers are grouped according to the relative importance placed on various benefits. It has been claimed to have greater potential for translating segmentation results into marketing strategy than other segmentation approaches (Young, Ott and Feigin 1978), and the segments obtained show a high degree of internal consistency and temporal stability (Calantone and Sawyer 1978).

Current benefit segmentation techniques commonly use a two-stage procedure (e.g. Hauser and Urban 1977, Currim 1981). At the first stage, benefit importances are obtained at the individual level, either by direct rating or by estimation. At the second stage, subjects are clustered on the basis of similarity of benefit importances.

When ideal-point models are used to analyze consumer preferences, ideal points are used as a basis for segmentation (Johnson 1971, Frank, Massy and Wind 1972, Dickson and Ginter 1987). Ideal points can be obtained either by direct rating (Johnson 1971) or by statistical inference (Frank, Massy and Wind 1972). The segmentation procedure employed here is also a two-stage one.

The traditional two-stage approaches to benefit segmentation have a number of important methodological limitations however.

First, the validity of segments obtained by a clustering procedure is highly dependent on the validity of the assessed idiosyncratic benefit importances. Direct ratings of benefit importances are easier to collect and can be obtained when the number of products is small, but they are in general less reliable than inferred weights (Slovic and Liechtenstein 1971, Fishbein and Ajzen 1975, Neslin 1981, Brookhouse et al. 1986), and may be biased as, for example, subjects may rate prominence, salience or relevance, rather than value importance (Calantone and Sawyer 1978, Wilkie and Cohen 1977). People tend to overestimate the importance of less important product attributes (Slovic and Liechtenstein 1971). Consequently, benefit segmentation on the basis of inferred weights is generally recommended (Urban and Hauser 1980).

However, the superior reliability of inferred weights reveals itself only when a sufficient number of product evaluations are available at the individual level. Evidence indicates that the relevant set (i.e., the product alternatives the consumer is familiar with) is typically rather small (Urban 1975, Silk and Urban 1978). This implies that, in practice, few degrees of freedom are available for estimation at the individual level, which may restrict the models that can be fitted to the simpler types, resulting in model misspecification, biased estimates, and poor predictions (Carmone and Green 1981). Multicollinearity of the predictors (caused by correlation between the perceptual dimensions in compositional analysis or highly fractionated designs in decompositional analysis) may lead to unreliable estimates (Wedel and Kistemaker 1989). Due to the small number of degrees of freedom the estimates are sensitive to the measurement error. As a result, idiosyncratic inferred weights are unreliable or even can not be estimated at all (e.g. Hauser and Urban 1977, Urban and Hauser 1980, Tybout and Hauser 1981). In these cases, it is questionable or impossible to group the respondents on the basis of similarity of estimated importance parameters.

Second, the criteria for grouping subjects with hierarchical or nonhierarchical methods do not maximize the accuracy with which overall product evaluations are predicted within segments (Kamakura 1988). One may obtain a 'good' cluster solution (in terms of the homogeneity of estimated importance weights), without any appreciable increase in predictive power over the unsegmented model. Predictive accuracy, however, is a key measure for evaluating market segmentation results and for developing a marketing strategy (Hauser and Urban 1977, Currim 1981).

In the segmentation literature a number of procedures have been developed that alleviate one or both of the above disadvantages. These will be discussed in the following sections.

6.2 Flexible and componential segmentation

Both flexible (Wind 1978) and componential (Green 1977) segmentation are closely connected to conjoint analysis. Both of these methods alleviate the disadvantages connected to clustering individually estimated part-worths, and focus on the prediction of segment response.

Componential segmentation (Green 1977, Green and DeSarbo 1979) involves an extension of conjoint analysis in that not only product profiles, but also respondent profiles are generated according to an experimental design. Respondents matching the profiles are sought from an available sampling frame. Each of the selected respondents is then administered the task of evaluating the product profiles. The respondents' evaluations are submitted to the componential segmentation model. This decomposes the response of respondent j (with levels $j_1 \ldots j_M$ of the M consumer descriptors, $m = 1 \ldots M$) to the k-th product profile (with levels $k_1 \ldots k_P$ of the P product characteristics, $p = 1 \ldots P$) into the part-worths of the levels of the product characteristics, and their interactions with the levels of the subject characteristics:

$$y_{k_1 \ldots k_P j_1 \ldots j_M} = \Sigma_p u_{k_p} + \Sigma_m v_{j_m} + \Sigma_p \Sigma_m w_{k_p j_m} \qquad (6.1)$$

where

u_{k_p} = the contribution due to level k_p of brand attribute p.

v_{j_m} = the contribution due to level j_m of consumer variable m.

$w_{k_p j_m}$ = the contribution due to the interaction of level k_p of brand attribute p with level j_m of person variable m.

The (least-squares) parameter estimates are obtained with a stage-wise fitting procedure (Green and DeSarbo 1979). Given these estimates, predictions can be made of the evaluation of any combination of product attributes in any respondent segment, defined by a combination of the consumer characteristics in the model.

Drawbacks of componential segmentation are that it requires the availability of a large database for consumer selection, that the consumers selected do not constitute a random sample, and that a limited number of segmentation variables have to be specified a priori. Practical use of the method requires cross-validation on hold-out data. The performance of componential segmentation depends upon the specific consumer characteristics selected in applications. With respect to predictive validity, Moore (1980) provided some empirical evidence for the superiority of the two-stage clustering approach over componential segmentation.

In flexible segmentation (Wind 1978), the results of a conjoint analysis study are fed to a computer program that simulates consumer choice behavior with respect to new products. The marketing manager interactively enters new product profiles into the simulation. These profiles are defined as combinations of the specific levels of the attributes included in the conjoint analysis study. The computer simulation determines consumers' share of choice of existing products, and the switching to the new product (based upon the assumption that consumers choose the brand with the highest utility). The market manager selects target segments on the basis of predicted switching behavior. The segments are profiled with demographic, life-style and other consumer characteristics using discriminant analysis.

Little is known about comparisons of the performance of flexible segmentation with other methods or the validity of the resulting segments. Flexible segmentation was specifically developed for marketing new products, and in fact entails a grouping of consumers on the basis of predicted behavioral intentions.

6.3 Approaches that maximize the predictive accuracy

Some of the major defects of the two-stage approach to benefit segmentation is the use of clustering approaches that do not maximize the predictive accuracy of behavior or behavioral intentions within segments. In relation to the responsiveness criterion for effective segmentation, it is just this ability of segmentation models to predict (the predisposition to) behavior that is of crucial importance for marketing strategy. Christal (1968), Bottenberg and Cristal (1978), Hagerty (1985), Ogawa (1987) and Kamakura (1988) developed procedures for maximizing predictive accuracy in market segmentation. These procedures will be described below.

The hierarchical clustering method "judgment analysis", JAN (Christal 1968, Bottenberg and Christal 1978, Lutz 1977), starts with single-subject clusters. In the procedure two previously defined clusters are combined successively in such a way that the loss in overall predictive efficiency (measured by R^2) of the regressions of preference on attributes within clusters is minimal. The regression model specification of JAN, at step s of the iterative process, contains the interactions of the P independent variables (x_{jkp}, $p = 1...P$; $k = 1...K$ denotes the products; $j = 1...n$ denotes the individuals) with the dummy variables indicating clusters ($z_{ij}^{(s)} = 1$ for j in cluster i, and 0 otherwise; $i = 1...c^{(s)}$; $c^{(s)}$ is the number of clusters at step s). At step 0 each individual j is considered a separate cluster. The model predicting preferences (y_{jk}) in step s is:

$$y_{jk} = \Sigma_p \Sigma_i \beta_{ip} x_{jkp} z_{ij}^{(s)} + \epsilon_{jk} \tag{6.2}$$

In the next step of the iteration ($s = s+1$) a new cluster dummy is calculated, say $z_{1j}^{(s+1)} = z_{1j}^{(s)} + z_{2j}^{(s)}$, and the multiple regression is run as a function of $z_{1j}^{(s+1)}$ and the remaining dummies. This is done for any combination of two dummies, and the combination giving a minimal decrease in R^2 is retained. The process is repeated at the following stages of the procedure. The authors indicate that the computational time can be reduced, because only the sum of squared errors (SSE) for the two groups being linked at each stage is to be computed. The method avoids clustering the instable individual estimates of the preference weights. However, it was shown by Adler and Kafry (1980) that JAN is identical to well known hierarchical clustering techniques, applied to preferences predicted with individual level models.

Within the context of conjoint analysis, Hagerty (1985) proposed a factor analytic approach that maximizes predictive accuracy. He assumes that a general weighting scheme transforms the K×n matrix of responses Y. (n denotes the number of subjects, K the number of stimuli and c the number of clusters.) The weighting scheme is denoted by the n×n matrix A ($A = S(S'S)^{-1}S'$, where S is a n×c matrix of cluster memberships. An expression for the expected predictive accuracy of this weighting scheme, with respect to validation stimuli, was developed. The optimal scheme for weighting consumer responses was derived by maximizing the predictive accuracy. The least-squares estimates of the (P×n) matrix of part-worths C, under optimal weighting, are:

$$C = (X'X)^{-1}X'YSS', \tag{6.3}$$

where X is the K×P matrix of predictors (the conjoint experimental design). The optimal weighting matrix S, consists of the first c eigenvectors of the correlation matrix between subjects, $1/(nK)(Y'Y)$. For interpretation of the resulting factor space, the idealized respondent and directional cosines method were suggested (Hagerty 1985). A problem associated with the optimal weighting method is in fact the interpretation of the factor solution in terms of segments. It has been shown (Stewart 1981) that the number of factors from a Q-type factor analysis is unrelated to the number of clusters present. The identification of clusters from the factor analysis solution is not straightforward; the approach suggested by Hagerty (1985), which uses optimal weighting in combination with cluster analysis, results in a loss of predictive accuracy.

Ogawa (1987) presented a simultaneous approach to estimation and segmentation in conjoint analysis. Based on a logit model, he proposed a ridge-regression-like procedure for estimating individual part-worths from profile rankings. (A sum of squared part-worths is added to the individual likelihood to be maximized, to ensure estimability at the individual level). The proposed aggregation procedure starts with single-subject clusters and is hierarchical. At each stage, previously defined clusters are combined in such a way that the aggregate log-likelihood is maximized. The method overcomes the weakness of unstable estimates of individual part-worths, at the initial stage of the algorithm, through the ridge-like procedure. At later stages the problem is alleviated by aggregating respondents. The individual estimates are biased due to the ridge procedure however, and because of hierarchical aggregation, misclassifications at earlier stages may carry on to higher levels. Aggregation on the basis of Euclidian distances of (ridge-like) estimates of cluster part-worths, appeared at least to equal the performance of the log-likelihood criterion, and was more attractive with respect to computational time.

Kamakura (1988) suggested a hierarchical least-squares procedure for segmentation in conjoint analysis. In Kamakura's procedure, benefit segments are formed in a such a way that the ability to explain each consumer's preferences with segment level part-worth estimates is maximized. Although collinearity leads to unreliable individual part-worth estimates, it does not affect the predicted responses, on which Kamakura's method is based. In the procedure, the preference functions for each of n individuals are expressed as:

$$\mathbf{Y} = \mathbf{XBG} + \mathbf{E} \tag{6.4}$$

where **Y** and **X** are defined as above, **B** is a P×c matrix containing the preference weights, **E** is a K×n matrix of random errors, and **G** is a c×n matrix defining non-overlapping cluster membership. The allocation of subjects in a predetermined number of segments is found using an agglomerative algorithm that maximizes the predictive accuracy index of Hagerty (1985). Kamakura demonstrated that maximizing the predictive accuracy for a fixed number of segments is identical to minimizing the overall SSE for the clustered solution. Whereas the two-stage procedure joins clusters with the highest sum of squared cross-products of part-worth estimates, Kamakura's procedure joins two clusters with the highest cross-validity of predicting the preferences for one cluster with the estimates of the other. (The two procedures are identical for orthogonal designs with only two-level attributes.) Despite the considerable number of pairwise linkages and regression function estimates, it was shown that the computations can be performed within reasonable time. A drawback related to the hierarchical algorithm is that misclassifications at earlier stages of the algorithm may carry on to higher levels.

6.4 Conclusions

As the approaches of Bottenberg and Christal (1978), Hagerty (1985), Ogawa (1987), and Kamakura (1988) group consumers into segments in such a way that the ability to predict segment behavior or behavioral intentions is maximized, they are appealing from both a marketing and a statistical point of view. The methods are all tailored to benefit segmentation, one of the most powerful approaches to segmentation.

The approaches produce a more valid segmentation than the traditional two-stage approach, because they are not based upon possibly unreliable individual estimates of importances or ideal points. The model chosen in each application will, however, at best provide a partial representation of each consumers individual decision process. Whereas conclusions based upon individual estimates of the importance weights may be severely affected by misspecification of the model, the predictions of misspecified compensatory models can be quite accurate (Carmone and Green 1981, Lynch 1985), even when the information processing strategy is complex and varies across consumers (Green and Srinivasan 1978). Segmentation approaches maximizing predictive accuracy thus do not rely heavily upon the psychological reality of the models, which are used only as a basis for predicting consumer behavior. Besides, these approaches minimize the discrepancy of measured and predicted purchase predisposition. Measures of purchase predisposition can be considered proxy variables of consumer demand, and clustering consumers on this basis is desirable from a normative perspective, as formulated by Dickson and Ginter (1987). Although their definition of the normative segmentation ideal is founded upon economic theory, their formulation of the demand function recognizes partial consumer information and is consistent with multi-attribute models of consumer behavior.

Yet, the procedures described suffer from a number of problems. They are based upon hierarchical clustering procedures or factor analysis. The hierarchical methods have been shown to be inferior to nonhierarchical clustering methods (Punj and Stewart 1983), and factor analysis has distinct disadvantages for grouping tasks (Stewart 1981). Models that are overparameterized at the individual level can not be dealt with and the methods yield nonoverlapping partitions. The validity of nonoverlapping approaches has been questioned, especially when usage situations are not taken into account.

PART 2

CLUSTERWISE REGRESSION:

DEVELOPMENTS AND APPLICATIONS

7 BENEFIT SEGMENTATION USING CLUSTERWISE REGRESSION

7.1 Introduction

In this chapter a method will be proposed that is tailored to benefit segmentation, and that overcomes some of the problems related to the traditional two-stage procedure described in the last chapter. The data for the analysis are assumed to consist of perceived product dimensions, obtained by factor analysis or multidimensional scaling, or of product profiles, resulting from experimentally varied attribute levels in conjoint analysis. Preferences are assessed for each individual in the sample (using rating scales, paired comparisons, or preference rankings.) A model is specified that relates preferences at the individual level to (functions of) product dimensions or profiles.

The disadvantages of the traditional two-stage procedure for benefit segmentation have been discussed in Chapter 6. When a consumer's relevant set of products is small, the x matrices in the individual regressions may not be of full rank, and the preference weights are not estimable. This problem can be solved by deleting a sufficient number of terms from the model. A serious disadvantage of this approach is that the estimates of the preference weights in the reduced model are biased if the model terms excluded are good predictors of preference. Even if the individual x matrices are of full rank, there are often few degrees of freedom for estimation because of small sets of relevant products, or limitations in design or data collection. This results in near-collinearity and unreliable estimates of the preference weights, as the estimates are then very sensitive to measurement error (Mason, Gunst and Webster 1975). Grouping individuals on the basis of estimated preference weights can then hardly be expected to identify existing segments.

In the present chapter a method is described that simultaneously estimates segments and preference functions within segments, and that alleviates the problem of unreliable individual estimates, since preference functions are estimated within segments across subjects and products, which provides more degrees of freedom. Preference models can even be fitted which would be severely overparameterized at the individual level, and misclassifications of subjects due to errors in estimates of the individual preference weights are avoided. Moreover, it will be shown that the procedure explicitly maximizes the predictive fit. It deals with replicated observations per subject, such as in preference data. The procedure is called clusterwise regression after Späth (1979, 1982).

7.2 The method

7.2.1 *Clusterwise regression*

Let:

$\mathbf{y}_j$ = the (K_j×1) vector of preferences of individual j for K_j products (j = 1...n),

$\mathbf{X}_j$ = the (K_j× P) model matrix for individual j, the P columns being functions of the product dimensions, depending on the model that is appropriate for the analysis,

$\mathbf{b}_j$ = the (P×1) vector of preference weights of individual j.

Consider an analysis of the preference data in which the following model is assumed:

$$\mathbf{y}_j = \mathbf{X}_j \mathbf{b}_j + \mathbf{e}_j, \qquad \mathbf{e}_j \sim N(0, \mathbf{I}\sigma_j^2), \qquad j = 1...n, \qquad (7.1)$$

where $\mathbf{e}_j$ is a vector of independently Normal distributed error terms [1]. Assume each of the $\mathbf{y}_j$ to arise from one of c segments, assume c to be known and assume the parameter vectors $\mathbf{b}_j$ of the individual preference models to be the same for the set of n_i individuals (denoted by C_i) in segment i ($\mathbf{b}_j = \mathbf{b}_i$, $j \in C_i$) [2]. Now let:

N_i = the number of observations in segment c, $N_i = \Sigma_j K_j$, $j \in C_i$,

$\mathbf{y}_i$ = the (N_i×1) partitioned vector of preferences of individuals in segment i, consisting of the n_i subvectors $\mathbf{y}_j$, $j \in C_i$,

$\mathbf{X}_i$ = the (N_i× P) partitioned model matrix, consisting of the n_i submatrices $\mathbf{X}_j$, $j \in C_i$,

$\mathbf{b}_i$ = the (P×1) vector of preference weights in segment i.

If $\sigma_j = \sigma_i$ for all $j \in C_i$, the model for the N_i observations in segment i is:

$$y_i = X_i b_i + e_i, \qquad e_i \sim N(0, I\sigma_i^2), \qquad i = 1 \ldots c, \tag{7.2}$$

If $P < N_i$ in each of the c segments, the X_i are of full-rank P and the b_i are estimable given the partition of subjects into c segments.

The objective is to find the partition of the subjects into the c segments. By analogy with Scott and Symons (1971), the maximum likelihood (ML) estimates of the (c×n) matrix U, containing the memberships, u_{ij} ($i = 1 \ldots c$, $j = 1 \ldots n$), of subject j in segment i, can be derived ($u_{ij} = 1$ for $j \in C_i$, and $u_{ij} = 0$ otherwise). Assuming the partition of subjects into c segments to be known, the log-likelihood function for the parameters b_i and σ_i is given by:

$$\ln(L) = -1/2\, \Sigma_i (y_i - X_i b_i)'(y_i - X_i b_i)/\sigma_i^2 - 1/2\, \Sigma_i N_i \ln(\sigma_i^2) \tag{7.3}$$

For each possible partition into c segments the likelihood is maximized by the ordinary least-squares estimates of b_i and σ_i, $\hat{b}_i$ and $\hat{\sigma}_i$. By substituting $\hat{b}_i$ and $\hat{\sigma}_i$, dividing by N ($N = \Sigma_j N_j$), taking the antilog, and dropping a constant term, exp(-K/2N), Eq. 7.3 can be reduced to:

$$L^* = \Pi_i \hat{\sigma}_i^{2(N_i/N)} \tag{7.4}$$

The maximum likelihood (ML) estimate of the membership matrix U, is that grouping of the n subjects into c segments which minimizes the criterion L^*, the weighted geometrical average of the within-segment error mean squares. If $\sigma_i = \sigma$ for all i, the ML estimate is the partition which minimizes $\hat{\sigma}^2$, the pooled residual mean square (RMS) of the regressions within the c groups. This is equivalent to maximizing the predictive accuracy, R^2, as for a fixed number of segments both the degrees of freedom and the total sum of squares are fixed. For $N_i = 1$ and $\sigma_i = \sigma$ the clusterwise regression problem of Späth (1979) minimizing the L_2 norm is obtained.

The partition minimizing the criterion L^* can be found by comparing all possible partitions of subjects into c segments (as was done by Scott and Symons in their application in 1971). For large numbers of subjects the computational time required would be excessive, and therefore the transfer algorithm of Banfield and Bassil (1977) can be used in practice to obtain the partition. This algorithm starts from a given classification of the individuals, and has two phases, one of transferring subjects from one segment to another, and one of swapping two subjects between segments. Each possible transfer and swap is tested, and executed if they reduce the value of L^*, until no further improvement can be realized. If swapping is successful, the transfer phase is reentered. The transfer phase of the Banfield and Bassil algorithm is identical to Späth's exchange method (1977). When the Banfield and Bassil algorithm is applied, transfers from clusters with P observations or less are not permitted. As with most divisive methods, optimal classifications found may not be unique, and may not be global optima. Banfield and Bassil suggest to use a classification with more than c clusters as a starting point and to work down to the desired number of c clusters to help avoiding local optima, or else to use different random starting classifications.

7.2.2 *Significance testing*

Although estimates of the parameters with clusterwise regression are ML estimates, the asymptotic properties do not apply, because the number of parameters estimated is always close to the number of observations. As the distribution of the minimum of $\hat{\sigma}$ is unknown, the usual t- and F-tests for the significance of the regressions within clusters are not valid, but simplified Monte Carlo significance tests can be used. In Monte Carlo test procedures the outcome of the test is determined by the rank of a statistic derived from the observed data, relative to the values of that statistic derived from random samples (the reference set). The reference set is generated in accordance with the hypothesis being tested. For the simplified Monte Carlo test the reference set consists of M-1 samples, and the null hypothesis is to be rejected if the test criterion from the observed data is greater (or smaller) than M-M(α/2) or more of the values from the reference set (α is the level of significance of the two-sided test, and both M and M(α/2) are integers). The power of the test increases with M, and approaches that of the uniformly most powerful test in the limit. An expression for the power is given by Hope (1968). For the clusterwise regression problem the samples of the reference set can be obtained by permuting the observed preference scores randomly among products for each subject.

7.2.3 *Computer program*

A FORTRAN computer program RMSCLUST was developed. This program starts from a random or (partially) pre-set classification and uses the Banfield and Bassil algorithm (1977) to minimize the criterion L^* (Eq. 7.4) either with or without the assumption of equal σ_i. In the calculations the program uses individual sums of squares and products (SSP) matrices of the y and x variables ($(1+P)\times(1+P)$ symmetric matrices), so that the number of products per subject does not affect the computations. Moreover, the calculation of the criterion value requires only an inversion of the within-cluster SSP matrices involved, at each transfer or swap being tested, whereas the calculation of a within-cluster SSP matrix involves a simple addition or subtraction of the ssp matrix of the individual added to or removed from that cluster. The algorithms described by Herraman (1968) and Clarke (1982) were used for scaling and inverting the SSP matrices. The iteration procedure does not require full regression models to be calculated, which increases computational efficiency. The RMSCLUST program can start from a given number of segments and automatically works down to a given final number of segments. The swap phase of the algorithm is optional, the transfer phase is similar to Späth's exchange algorithm (1977, 1979, 1981, 1982, 1986). The program output includes monitoring of the numbers of transfers and swaps tried and executed as well as the changes in the criterion value during the iteration process. An option for generating reference sets for a Monte Carlo permutation test is included (Kistemaker and Wedel 1988).

7.2.4 *Performance on synthetic data*

To investigate the performance of the algorithm in various situations, a simulation study was conducted in which the method was applied to synthetic data (see e.g. DeSarbo 1982). Both the x variables and the coefficients (preference weights) were randomly generated from a common uniform distribution for each segment, using the pseudo-random number generator implemented in the statistical package GENSTAT (Alvey et al. 1977). The y variable was calculated from the x variables without error. Four factors were varied, according to a half replicate of a 3^4 factorial design (27 trials, Cochran and Cox 1957): 1) the number of x variables (1, 2, or 4); 2) the number of segments (2, 3, or 4); 3) the number of products (1, 2, or 6); 4) the number of subjects per segment (5, 10, or 25).

Each trial was analyzed 6 times with a different starting partition. The percentage of times the actual segments were perfectly recovered was registered, as well as the numbers of transfers and swaps executed and the CPU time required (on a VAX11/750 computer). These measures of cluster recovery and computational performance were analyzed by analysis of variance for main effects. In one trial, with 4 x variables and segments of 5 subjects, the algorithm could not be applied due to collinearity. Table 7.1 shows the results.

The recovery was satisfactory, 72.8% of the true segments being perfectly recovered by the algorithm. However, the algorithm fails to identify segments more frequently with more x variables, more segments, fewer products and fewer subjects per segment. This means that the quality of the solution derived depends on the number of parameters to be estimated and the size of the sample. As is generally the case in nonlinear estimation problems, parameter recovery declines as the number of parameters to be estimated increases. This is a potential problem associated with the algorithm, which is most likely caused by convergence to local optima. It was observed that in many instances where the true cluster structure is not fully recovered, one or more of the true segments are split into two or more segments in the clusterwise regression solution, while other true segments are merged, because the number of segments is fixed. These problems of local optima can be solved in practice by having the algorithm started from a number of different starting partitions, or with a number of segments larger than c and work down to the desired c clusters (a procedure not employed in this Monte Carlo experiment). Increasing the size of the sample in terms of products or subjects improves segment/parameter recovery, a common result in statistical estimation.

DeSarbo and Cron (1988) report similar results for the effect of data characteristics on parameter recovery in clusterwise regression.

Table 7.1
Results of the simulation study on the effects of four factors on clusterwise regression performance

Factor[1]	Level	Perc. of segments recovered	Number of transfers executed	Number of swaps executed	CPU (s)
x Variables					
	1	85.2^a	29.4^a	0.26	23.4^a
	2	74.1^a	30.9^a	0.63	31.3^a
	4	59.3^b	48.1^b	0.58	108.5^b
Segments					
	2	88.9^a	11.2^a	0.24^a	11.2^a
	3	79.6^a	35.2^b	0.17^a	14.6^a
	4	50.0^b	62.0^c	1.06^b	110.5^b
Products per subject					
	1	59.3^a	50.1^a	1.21^a	82.1^a
	2	68.5^a	30.6^b	0.26^b	37.5^b
	6	90.7^b	27.7^c	0.00^b	43.7^b
Subjects per segment					
	5	66.7^a	12.7^a	0.87^a	19.7^a
	10	59.3^a	35.9^b	0.28^b	60.8^b
	25	92.6^b	59.8^c	0.31^b	82.8^c

[1] Factor level means sharing a superscript are not significantly different at $p < 0.05$.

Regarding computational performance, the number of transfers executed increases with an increasing number of x variables, segments, and subjects, and with a decreasing number of products (the number of transfers tried shows the same pattern, but is about 10-15 times higher). The numbers of swaps executed increases with increasing numbers of clusters, and with decreasing numbers of subjects and products. Since the number of swaps tried (which shows a similar pattern) is high (1100 on the average), and the number of swaps executed is small

(0.5 on the average), it is important to note that the swapping phase of the algorithm is efficient for few products, few subjects and large numbers of segments, irrespective of the number of x variables. Discarding the swapping phase in other instances will reduce the computational time required. The computational time shows a pattern consistent with that of the number of transfers and swaps. Summarizing, the computational effort required increases both with the number of parameters to be estimated and with the size of the sample. On the average, the process took 53 seconds CPU time.

7.3 Application to data of meat preference

7.3.1 *Data*

The increasing number of elderly people in most European countries, and problems connected with aging, such as lack of mobility, unhealthiness, bad dentition, and physical disability, make the elderly a target of interest, especially for the food industry. Nutritional guidelines, which agree in that they recommend a reduction of fat intake, affect food marketing both in the USA and in European countries (Richardson 1987). The elderly have been designated by public authorities as a group worthy of particular attention. Meat, as a major contributor to fat intake, has been pinpointed as a food group to be reduced in the diet.

To study the factors influencing meat preference among elderly people, in a nationwide random sample of 199 subjects aged 65 to 80, data were collected on preferences and perceived product characteristics with respect to meat products (Wedel, Hulshof and Löwik 1986). Subjects were asked to rank photographs of 11 raw meat products in order of their preference (ties were permitted). Preference values ranged from 1, corresponding to the lowest preference, to 11, corresponding to the highest preference. Twenty attributes were evaluated for each product and reduced to four perceptual dimensions, using factor analysis (Hauser and Koppelman 1979), explaining 16, 14, 13 and 8% of the variance of the attribute ratings respectively. These dimensions, labeled quality, fatness, exclusiveness, and convenience, were to be related to preferences, by linear regression, to obtain estimates of the preference weights[3]. Average factor scores for the 11 meat products are presented in Table 7.2.

Table 7.2
Averages of perceived sensory quality (SE), fatness/unwholesomeness (FA), exclusiveness (EX), and convenience (CO), for the 11 types of meat

	SE	FA	EX	CO
Thin pork steaks	-0.18	-0.20	0.58	0.64
Pork steaks	0.27	-0.21	-0.05	0.08
Pork loin/rib chops	0.36	-0.12	0.47	0.23
Pork shoulder chops	0.09	0.34	-0.49	0.05
Minced meat	-0.06	0.07	-0.82	-0.04
Pork belly steaks	-0.05	0.75	-0.80	0.58
Pork sausages	-0.28	0.50	-0.70	0.35
Sirloin steaks	-0.22	-0.69	1.01	1.06
Lean beef steaks	-0.07	-0.76	0.21	-1.12
Rolled pork	0.24	0.06	0.92	-0.87
Brisket beef steaks	0.09	0.27	-0.32	-0.96

Because rank order preferences were collected, the errors may not be independently Normal-distributed. Examination of the residuals of the total sample regression of preference on the perceptual dimensions (Table 7.3) showed that the residuals have a unimodal distribution with 64.4% between ± SD and 96.5% between ± 2SD. Tests for skewness (0.18; SE=0.05) and curtosis (-0.57; SE=0.10) indicated a slight positive skewness and a distribution somewhat flatter than Normal. Both the serial correlation coefficient (0.123) and the Durbin Watson statistic (1.75) indicate a weak positive serial correlation of the residuals. The Normal probability plot showed no significant departure from linearity, and we conclude that for our purposes the distribution of the residuals can be sufficiently approximated by a Normal. Even when the assumptions of independence and constant variance of the errors do not hold approximately (the only assumptions necessary in very large samples for the estimates to have ML properties), least squares can be and have been used (Hauser and Urban 1977) as an estimation procedure. Moreover, these assumptions are not necessary for significance testing, because the Monte Carlo procedure is used, which does not depend on Normality assumptions, or ML properties.

7.3.2 *Consumer segmentation*

In the application the number of subjects n was 199, the number of products K_j was 11 for all j, the number of parameters P was 5, and the total number of observations N was 2189.

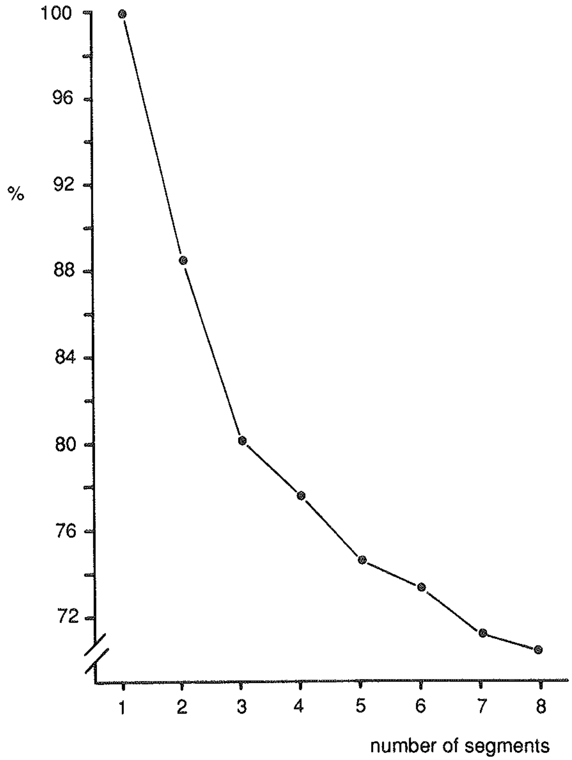

Figure 7.1. Plot of $1\text{-}R_a^2$ (as a percentage of the unsegmented solution) against the number of segments for the meat-data

The algorithm for clusterwise regression was started with a random classification of $c=8$ segments and worked down to 2 final segments, minimizing the RMS. 50,346 Transfers were tried and 853 executed, 122,627 swaps were tried and only 3 executed.

The process required 7234 seconds CPU time on a VAX 11/750 computer; a more efficient initial classification based, for example, on a clustering of preferences probably would reduce the CPU time required. Figure 7.1 shows a plot of $1\text{-}R_a^2$ (the adjusted R^2; Theil 1971), expressed as a percentage of the total sample value, against the number of segments.

The statistics did not show an optimum within the numbers of segments investigated.

The three-segment solution was subjectively selected on the basis of an apparent elbow in the curves in Figure 7.1. This solution had a RMS of 5.019. Subsequently, clusterwise regression was applied 25 times with different random starting classifications ($c=3$). Seventeen times a solution with a RMS of 5.019 was found, eight times the solution was identical to the one described above, nine times a different solution was obtained in which only 2 subjects were classified differently. Eight times a solution with a RMS of 5.020 was found in which six subjects were in different classes than in the 'optimal' solution.

Consequently, the taste variations for the meat products seem not to be densely clustered, and the optimum found does not seem to be local, but appears to be a rather flat global optimum. For the clusterwise regression solution $R_a^2=0.485$, whereas for the total sample $R_a^2=0.362$.

Table 7.3 shows the preference weights estimated within the clusters obtained. To evaluate their significance 13 data sets were generated by permutation of the preferences within individuals, and 39 values of the t-statistic were obtained for each dimension (13 times 3 segments). The 2.5th and 97.5th percentiles of the distribution of the t-values of the reference set are given in Table 7.3.

Table 7.3
Results of the three-segment solution of the clusterwise regression analysis of data on meat preference

	Segment 1 n=69	Segment 2 n=82	Segment 3 n=48	Total sample n=199
Constant	5.87[b]	5.64[b]	7.39[b]	5.83
(47.43;66.13)[a]	(70.87)	(74.43)	(70.93)	(109.25)
Quality	1.88[b]	2.20[b]	1.42[b]	1.73
(-9.03; 7.22)	(21.29)	(26.44)	(16.19)	(32.35)
Fatness	0.03	-0.81[b]	-0.63	-0.43
(-7.79; 7.20)	(0.39)	(-10.85)	(-6.46)	(8.08)
Exclusiveness	-0.92[b]	0.53	0.23	-0.06
(-9.95; 9.40)	(-11.48)	(7.22)	(2.15)	(1.11)
Convenience	0.55	0.59	0.45	0.60
(-10.27; 8.74)	(6.73)	(7.90)	(4.64)	(11.25)
R_a^2	0.488	0.485	0.453	0.361

[a] The 2.5th and 97.5th percentiles of the distribution of t-values in the reference set are given in parenthesis.
[b] Significant at the 5% level according to the Monte Carlo test procedure.

The coefficients of sensory quality in all segments, of fatness in segment 2, and of exclusiveness in segment 1 are significant at the 5% level. The Monte Carlo test procedure is seen to be much more conservative than the inappropriate t-test, which would have indicated all but one of the estimated coefficients to be significant. The segmentation analysis revealed that preference was most strongly related to perceived sensory quality in all segments. In fact, elderly people in segment 3 (24%) primarily evaluated meat in terms of sensory appeal. In segment 1 (35%) sensory quality was weighed against exclusiveness (negative), in segment 2 (41%) it was weighed against fatness. Convenience was consistently positively, but not significantly, related to preference in all segments. The three segments were described by a number of consumer characteristics, established in interviews: region (north, east, south, west, the large cities Amsterdam, Rotterdam or The

Hague), residence (urban, rural, large cities), age, sex, socioeconomic status (high, middle, low), income (minimum as defined by law, above minimum), civil status (single, married, widow(er)), diet (fat/energy-constrained, no diet), and place of purchase (butcher's shop, supermarket, elsewhere). Degree of self-care, mobility, masticatory function, and health status were classified by a physician as good or impaired. Nutritional knowledge was assessed by 30 questions about the nutrient content of products, to be answered as true or false, the score being calculated by adding correct responses and subtracting wrong ones. Cronbach's internal reliability coefficient was 0.71 for this test, while factor analysis for the examination of construct validity revealed that the first factor extracted had negative loadings for all but one of the test items, and explained 13.8% of the variance (Cronbach 1971).

Using chi-square tests for categorical and one-way ANOVA for interval data, region ($p<0.05$), residence ($p<0.10$), sex ($p<0.10$) and nutritional knowledge ($p<0.01$) were found to be related to cluster membership (Table 7.4). In a multivariate (log-linear) analysis of these variables only the association of nutritional knowledge with cluster membership remained significant ($p<0.05$). Discriminant analysis showed a 40.7% correct jackknifed classification of subjects into the three segments (11.6, 62.2 and 45.8% respectively) on the basis of the nutritional knowledge score. In general, the association of cluster membership with consumer characteristics was weak.

In segment 1, preferences for pork shoulder chops, sausages, minced meat, brisket steaks and pork belly steaks are high (Table 7.4). These products compete with respect to perceived sensory quality and low perceived exclusiveness (related to perceived price), in a segment consisting of somewhat more males, less subjects in the large cities and the western part of the country, and more in the other regions. Nutritional knowledge was intermediate (Table 7.4). Given the low perceived sensory quality of a number of the products in question, especially of pork sausages (Table 7.2), product modification to increase sensory appeal, without affecting price, would increase consumer preferences in this segment.

Table 7.4
Average preferences and consumer characteristics, of the three-segment solution of the cluster-wise regression analysis of meat preference[1)]

	Segment 1	Segment 2	Segment 3	Total sample
Average preferences				
Thin pork steaks	3.90[a]	5.93[b]	6.31[b]	5.32
Pork steaks	6.62	6.43	6.98	6.63
Pork loin/rib chops	7.29[a]	7.54[a]	8.75[b]	7.74
Pork shoulder chops	8.17[a]	5.71[b]	6.13[b]	6.66
Minced meat	6.96[a]	5.23[b]	6.21[ab]	6.07
Pork belly steaks	7.81[a]	4.17[b]	5.56[c]	5.77
Pork sausages	6.52[a]	4.54[b]	5.48[b]	5.45
Sirloin steaks	2.70[a]	6.94[b]	7.52[b]	5.61
Lean beef steaks	6.17[a]	8.10[b]	8.63[b]	7.56
Rolled pork	4.45[a]	5.72[b]	6.88[c]	5.56
Brisket beef steaks	6.16[a]	4.93[b]	5.69[ab]	5.54
Consumer characteristics				
Region (%)				
North	21.7	17.1	6.3	16.1
South	30.4	26.8	22.9	27.1
East	34.8	29.3	27.1	30.7
West	10.1	15.9	29.2	17.1
Large cities (%)	2.9	11.0	14.6	9.0
Residence (%)				
Urban	46.4	53.7	39.6	47.7
Rural	50.7	35.4	45.8	43.8
Sex (%)				
Male	55.1	36.6	45.8	45.2
Female	44.9	63.4	54.2	54.8
Nutritional knowledge	17.9[a]	19.7[b]	15.1[c]	18.0

[1] Cluster means sharing a superscript are not significantly different at $p < 0.05$.

In segment 2, preferences for pork loin/rib chops and sirloin and lean beef steaks are high. These products compete with respect to sensory quality and low perceived fatness, within a segment comprising more females, more subjects in the urban regions and the large cities and subjects with more knowledge of nutrition. Information on fat content, for through product or shelf labeling, required by consumers to make informed choices, as well as reduction of fat content of meat without affecting its sensory appeal, would increase consumer preferences in this segment. The increasing concern among consumers about diet and health is well documented (Breidenstein 1988, Baron 1988, Woodward 1988), and firms are already developing "light" varieties of certain types of meat.

In segment 3, preferences for the meat products are generally higher than in the other two segments but, again, lean beef steaks, sirloin steaks and pork loin/rib chops compete with respect to perceived sensory quality. A number of products, such as pork belly steaks, sausages and brisket beef steaks, are to a certain extent avoided by the consumers in this segment because of low sensory appeal. This segment has less knowledge of nutrition and consists of more people in the west as well as of the large cities, while relatively few subjects of the north belong to this segment.

7.3.3 *Comparison with the two-stage procedure*

In order to compare the clusterwise regression procedure with the two-stage procedure, preference weights were estimated for each individual, and subsequently used as a basis for segmentation. A clustering algorithm was applied minimizing the determinant of the pooled within-group covariance matrix. This method yields ML estimates of the memberships of subjects, given the number of segments and assuming the preference weights to follow a multivariate Normal distribution with identical covariance matrices within segments (Scott and Symons 1971). The method has been implemented in the statistical package GENSTAT (Alvey et al. 1977), and also uses the Banfield and Bassil algorithm. The partition obtained with clusterwise regression was used as a starting classification.

The percentage of subjects remaining in the initial solution was 60 (77, 68, and 25 for clusters 1 to 3 respectively). On the average, there was no difference in magnitude of the standard deviations of coefficients within segments between the two methods, whereas the clusterwise regression solution had a greater predictive efficiency: $R_a^2 = 0.411$ for the two-stage procedure, $R_a^2 = 0.485$ for the clusterwise regression solution. The averages for predicting the first and second preferences were 8.1 and 7.7 respectively for the two-stage procedure, and 8.3 and 7.9 respectively for clusterwise regression, while the averages for the last and second last preferences were 3.9 and 4.3, versus 3.2 and 3.9 respectively. The two-stage procedure predicted the first preference correctly for 43.3% of the subjects, the clusterwise regression procedure for 47.9% (these numbers can be compared to a random fraction of 18.2%). So, on the average, clusterwise regression predicts preferences somewhat better than the two-stage procedure. (Note that there were 5 residual degrees of freedom at the individual level and that the two-stage procedure will perform worse for smaller numbers.)

7.3.4 *Market structuring*

The method described can also be used to group products. Such an application could be useful in the context of market structure analysis to obtain groups of products with homogeneous preference functions and to determine competitive market structures. The analyses used in market structuring often parallel the two-stage segmentation procedure described above (Grover and Srinivasan 1987). The clusterwise regression algorithm was applied to the data on meat preference in order to cluster products, by using formulae 1 to 4 with the subscript $j = 1...n$ now denoting products, i the product clusters and C_i the set of products in cluster i. The algorithm was started with 11 single-product clusters. The results will be mentioned just briefly. Although a 7-cluster solution appeared to be optimal, the differences in R_a^2 were small and the 4-cluster solution was chosen on the basis of an apparent elbow in the curve ($K=11$, 0.410; $K=10$, 0.412; $K=9$, 0.412; $K=8$, 0.412; $K=7$, 0.412; $K=6$, 0.411; $K=5$, 0.409; $K=4$, 0.407; $K=3$, 0.398; $K=2$, 0.388).

The four clusters were as follows. Cluster 1: thin pork steak, pork steak, minced meat, and pork sausages. Preference weights were (t-values in parenthesis): sensory quality 1.5 (18.1), fatness -0.4 (-4.1), exclusiveness 0.0 (0.1) and convenience 0.8 (8.2). Cluster 2: pork rib and shoulder chops, pork belly steaks and lean beef steaks. Preference weights were: sensory quality 1.9 (21.9), fatness -0.5 (-5.8), exclusiveness 1.5 (1.6), and convenience 0.6 (6.1). Cluster 3 comprised sirloin steaks only. Preference weights were: sensory quality 2.7 (14.0), fatness -0.8 (-2.0), exclusiveness -0.6 (-1.9), and convenience 1.6 (4.9). Cluster 4 comprised rolled pork and brisket beef steaks, with the following preference weights: sensory quality 1.4 (10.5), fatness -0.4 (-3.2), exclusiveness 0.0 (0.0), and convenience 0.9 (5.1). By combining the structure of consumer perceptions (Table 7.2) with the results of this analysis it can be assessed which products compete within clusters, and indications are provided which desirable characteristics of products can be promoted, and which undesirable characteristics can be eliminated to improve consumer preference.

7.4 Concluding remarks

7.4.1 *Assumptions involved*

The assumption that coefficients for subjects within the same segment are identical might be somewhat restrictive in practice. If taste variation is also present within segments, the preference parameters within each segment can be assumed to follow a multivariate Normal distribution. Clusterwise regression can be applied and least-squares estimates of the mean preference parameters and the error variance can be shown to be unbiased, although they are not ML estimates.

The assumption of independently Normal-distributed error terms in the preference models is formally not correct when rank-ordered preferences are the dependent variable. As noted before, the relevant sets of alternatives should be identical across subjects. The estimates should be obtained with multinomial logit regression within clusters, and the segmentation should be performed by transferring and swapping individuals between clusters, using the log-likelihood as a criterion.

Clusterwise regression requires that the number of segments is known and that the segments do not overlap. In practice, the number of segments is determined empirically. In our example the optimum classification appeared to be flat. Some subjects could be transferred between segments without the criterion being changed dramatically. In such instances subjects may partially belong to more than one cluster, and the model oversimplifies the picture by forcing them into nonoverlapping segments. A fuzzy clustering method may be more appropriate in these situations. Such a method will be presented in Chapter 8.

The model applied assumes homogeneity of the preference functions across products. In practice, preference functions may vary among products, as was seen in the example when products were grouped by clusterwise regression. A clustering of products may be combined with segmentation analysis (Chapter 9).

7.4.2 *Related methods*

Späth (1979, 1982) originally proposed the method of clusterwise linear regression, which finds a given number of clusters of observations, such that the overall error sum of squares of the regression within clusters is minimal. For one observation per subject, the method is equivalent to the method described in this chapter. The simulated annealing method for clusterwise regression of DeSarbo, Oliver and Rangaswamy (1989) is also nonhierarchical, and differs from the present one mainly in the algorithm used. The simulated annealing procedure starts from a random initial partition, and iteratively specifies steps in a random direction in the parameter space. The new solution is accepted if it improves the criterion; if not, it is rejected with a probability proportional to the increase of the criterion value. Simulated annealing was devised as a method of optimization, less burdened with convergence to local optima, but the computational effort is reported to be considerable. The Banfield and Bassil algorithm can be considered as an extension of the exchange method described by Späth (1977, 1982). However, since the number of swaps executed in the application to real and synthetic data is relatively small, this second phase of the algorithm can be discarded in practice when the number of subjects and products is large. This probably reduces cluster recovery, but increases computational efficiency drastically, because the number of swaps tried is quite large. Unlike the simulated annealing procedure, the present method does not require the estimation of full regression equations at each iteration.

The procedure of DeSarbo, Oliver and Rangaswamy (1989) includes more than one dependent variable, as well as the option of overlapping clusters[4)]. The present method, however, includes the analysis of data with unequal numbers of products, and the analysis of data without the assumption of equal segment variances.

The ML methodology for fuzzy clusterwise regression of DeSarbo and Cron (1988) yields a fuzzy partition of observations, which are assigned to segments via estimated posterior probabilities. The method converges in few (less than a hundred) iterations in most cases, but information on computer time required is lacking, and the method does not deal with replicated measurements on subjects, such as in preference data. A full comparison of the present method with the methods of DeSarbo, Oliver and Rangaswamy (1989), and DeSarbo and Cron (1988), with respect to both the ability to identify global optima and computer time required, is left for further study.

7.4.3 *Conclusions*

We conclude that clusterwise regression for benefit segmentation is suited for finding consumer segments especially if collinearity plays a role in fitting preference models at the individual level. For problems of moderate size, the computational demands are not excessive. In applications, special attention should be paid to the problem of convergence to local optima and significance testing of estimated coefficients.

8 A FUZZY CLUSTERWISE REGRESSION METHOD FOR BENEFIT SEGMENTATION

8.1 Introduction

The clusterwise regression method described in the previous chapter alleviated two problems of the two-stage procedure: the estimates of the preference parameters being often unreliable because of low number of degrees of freedom at the individual level, and the criteria for grouping the estimates not maximizing the predictive accuracy. In the method segments were constrained to be mutually exclusive. The assumption underlying nonoverlapping segments is that segments are internally homogeneous and that subjects within a segment are distinct from those in other segments. The literature suggests however that the assumption of nonoverlapping clusters is often not met in practice (Inglis and Johnson 1970, Arabie 1977, Shepard and Arabie 1979, Hagerty 1985, Oppedijk van Veen and Verhallen 1986). Nonoverlapping clusters have been called "conceptually questionable" (Arabie et al. 1981, p. 310). Consumers may well belong to multiple segments when they desire several benefits from a product, possibly in relation to different consumption situations. Miller and Ginter (1979) showed that importance weights may vary considerably across usage situations. The negligence of consideration of usage context is equivalent to aggregation over contexts, differences among which will result in an increased of heterogeneity of the segments obtained, and an increase of the unexplained variance of the models. Consequently, there is a risk of oversimplification and loss of explanatory power when clusters are assumed to be mutually exclusive in benefit segmentation. The application presented in Chapter 7 was restricted to one usage situation, as this was expected to reduce overlap between segments that might have resulted from different preferences in different usage situations. Nevertheless, the solution did not change substantially when some subjects were transferred between segments, and it was concluded that these subjects might have belonged to more than one segment.

In this chapter, a fuzzy method for clusterwise regression (FCR) will be described, which was developed to solve the disadvantages of the two-stage approach outlined above, and specifically allows subjects to belong to more than one segment. It can be applied to the three-way data usually encountered in analyses of consumer preference: perceived product dimensions or product profiles (in decompositional analysis), to be related to preferences across consumers and products. The method yields a partition of these three-way data along the consumer mode, and estimates the regression of preference on product dimensions in a prespecified number of fuzzy clusters. At the same time the membership values of subjects with respect to these clusters are estimated. Insight into the pattern of competition of products within segments with respect to the relevant product dimensions is obtained, and the results permit marketing strategies to be developed, and to be targeted at the segments. Section 8.2 describes the proposed method, FCR, and compares it with other recently developed methods for benefit segmentation.

In 8.3, a Monte Carlo analysis of the performance of FCR on synthetic data is described. Section 8.4 reports the results of empirical comparisons of FCR. It is compared with the clusterwise regression method of DeSarbo et al. (1989) that accommodates for overlapping clusters, to investigate convergent validity, and with Hagerty's (1985) optimal weighting method for conjoint analysis, to investigate predictive validity. Section 8.5 presents a validation study which investigates both the cross-validity of the cluster solution and the preference prediction provided by FCR. Section 8.6 contains two empirical applications of FCR. In 8.6.1, FCR is used to analyze data on preferences for meat products, in 8.6.2, FCR is applied to study the different bases of the image of outlets selling meat. Section 8.7 contains the conclusions.

8.2 The method

8.2.1 *Fuzzy clusterwise regression*

The data for the analysis are assumed to consist of preferences of n subjects for K products, where the set C_j evaluated by consumer j (j=1...n) consists of K_j products. The preferences are to be related to P (functions of) perceived product dimensions or product profiles, and the importance weights are to be estimated[1]. Assume that the importance weights differ between a number of clusters, and that the number of clusters, c, is known. Assume that all subjects may have partial membership in all clusters. For cluster i (i = 1...c) the model relating the theoretical values of the preferences to product dimensions is:

$$\mathbf{h}_i = \mathbf{X}\mathbf{b}_i, \tag{8.1}$$

where

$\mathbf{h}_i$ = the (N×1) partitioned vector of theoretical preferences in cluster i ($N = \Sigma_j K_j$) consisting of the n (K_j×1) subvectors $\mathbf{h}_{ij}$ of theoretical preferences of subject j,
$\mathbf{X}$ = the (N×P) model matrix partitioned accordingly.
$\mathbf{b}_i$ = the (P×1) vector of importance weights within segment i.
$\mathbf{e}_i$ = $\mathbf{y} - \mathbf{h}_i$, is a (N×1) vector of estimated independent error terms[2)], e_{ijk}, where $k = 1...K_j$, and
$\mathbf{y}$ = the (N×1) vector of observed preferences.

The following development of the method is analogous to the the fuzzy c-varieties clustering algorithms (Dunn 1974, Bezdek et al. 1981a, b).
We let $\mathbf{U}$ denote a real (c×n) matrix with elements u_{ij} ($0 \le u_{ij} \le 1$):

$$\Sigma_i u_{ij} = 1; \qquad \Sigma_j u_{ij} > 0; \tag{8.2}$$

The matrix $\mathbf{U}$ gives a fuzzy c-partition of subjects in the sample, where $2 \le c < n$. The elements u_{ij} are the membership parameters, indicating the degree of membership of subject j in cluster i. The purpose is to estimate the c-partition $\mathbf{U}$, and the parameter vectors $\mathbf{b}_i$ (i = 1...c). Like Bezdek et al. (1981a,b), we define a weighted sum-of-squared-error criterion, representing the (weighted) sum of distances of subjects and products from the regression equations in all clusters:

$$J_{Rm} = \Sigma_i \Sigma_j \Sigma_k u_{ij}^{m} \underline{e}_{ijk}^{2}, \tag{8.3}$$

where the summations are across the appropriate values. The (user-defined) fuzzy weight parameter m, the exponent of u_{ij}, is a fixed real number, influencing the extent to which subjects belong to more than one cluster. The FCR algorithm follows from the necessary conditions for minimizing J_{Rm}, given c and m > 1, so that for all i and j:

$$\hat{\mathbf{b}}_i = (\mathbf{X}'\hat{\mathbf{U}}_i^m\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{U}}_i^m\mathbf{y}, \tag{8.4}$$

$$u_{ij} = 1/\Sigma_h(\hat{D}_{ij}/\hat{D}_{hj})^{1/(m-1)} \tag{8.5}$$

where

$$\hat{D}_{ij} = \Sigma_k \hat{e}^2_{ijk}, \tag{8.6}$$

Analogous to the geometrical interpretation of ordinary least squares estimation (Draper and Smith 1976), $\hat{D}_{ij}$ represents the squared distance of the preferences of subject j from the preferences predicted by the model in cluster i. $\hat{\mathbf{U}}_i$ is the diagonal matrix with estimates of the subject membership parameters with respect to cluster i:

$$\hat{\mathbf{U}}_i^m = \mathrm{Diag}(\hat{\mathbf{U}}_{ij}^m), \qquad \hat{\mathbf{U}}_{ij}^m = \hat{u}_{ij}^m \times \mathbf{I}_{K_j}, \tag{8.7}$$

where $\mathbf{I}_{K_j}$ is a $(K_j \times K_j)$ identity matrix. The proof is similar to that in Bezdek et al. (1981a) and is obtained by forming the Lagrangian of J_{Rm} with respect to the sum constraint on u_{ij}, and setting the derivatives with respect to **U** and $\mathbf{b}_i$ equal to zero (see Appendix A). Bezdek et al. (1981a) proved that estimators of the form (8.5) are necessary and sufficient for a local minimum of a fuzzy objective function of the form (8.3).

8.2.2 *The algorithm*

The FCR algorithm consists of a Picard iteration of Eq. 8.4 and 8.5 to an approximate solution (see Bezdek 1981a).

When for a subject the distances to more than one cluster are exactly equal to zero, the memberships in these clusters are equal to one, the constraints (Eq. 8.2) are not satisfied, and a tie-breaking rule is needed to determine cluster memberships. The case that for $j=j'$ at least one of the $\hat{D}_{ij'} = 0$ is taken care of by a tie-breaking rule similar to that given in Bezdek et al. (1981a). Let:

$$S = \{1,2\ldots c\},$$
$$S_{j'} = \{i \in S \mid \hat{D}_{ij'} = 0\},$$
$$n_{S_{j'}} = \text{the number of elements in } S_{j'}.$$

Then the tie-breaking rule is:

$$\hat{u}_{ij'} = 0 \text{ for } i \in S\text{-}S_{j'} \text{ and } \hat{u}_{ij'} = 1/n_{S_{j'}} \text{ for } i \in S_{j'}, \tag{8.8}$$

The proposed algorithm consists of the following steps:

1. At the first step of the iteration ($z=0$), initialize by
   a) fixing $2 \leq c < n$, and $m > 1$.
   b) selecting starting matrix $\hat{\mathbf{U}}^{(0)}$.
2. For $i=1\ldots c$, compute the $\hat{\mathbf{b}}_i^{(z+1)}$, using Eq. 8.4.
3. Calculate $\hat{\mathbf{U}}^{(z+1)}$, using Eq. 8.5 and Eq. 8.6. The tie-breaking rule (Eq. 8.8) is used for $j=j'$, if at least one $\hat{D}_{ij'}=0$.
4. Iterate between steps 2 and 3, and stop when the change in J_{Rm} is below a prespecified value (alternatively, changes in $\hat{\mathbf{b}}_i$ or $\hat{\mathbf{U}}$ can be used as a stopping criterion).

A potential problem related to the algorithm is convergence to local optima. This can be overcome by having the algorithm started from a number of random partitions, or by starting with a larger number of clusters to work down to the desired number of clusters (Banfield and Bassil 1977).

8.2.3 *Choosing m and c*

Selection of m

The parameter m provides flexibility with regard to the degree of overlap to be obtained. The value of m has to be fixed judgmentally however. The problem of selecting m is not specific to FCR, but common to the class of clustering algorithms from which it was derived (Bezdek et al. 1981a, b). It can be seen from Eq. 8.5 that values of m close to 1 will result in a (near) hard partition with all memberships close to 0 or 1. High values will lead to excessive overlap with all memberships close to 1/c, where c is the number of clusters on which the selection of m is based. Both types of solution are undesirable (Arabie et al. 1981) and have a near-zero variance of those memberships that indicate substantial cluster membership (indicated by values exceeding 1/c), while an optimum may be reached at intermediate values.

This suggests the pooled within-cluster membership standard deviation (SE_u), calculated across those subjects in segment i with $u_{ij} > 1/c$ ($i = 1 \ldots c$), as an heuristic criterion for the selection of m. Selecting the value of m which has the optimal value of this criterion guards against excessively overlapping and nonoverlapping solutions, both of which have near-zero values of SE_u[3].

In the section on the Monte Carlo analysis of FCR performance (8.3) the effect of m on parameter recovery will be evaluated, while in the empirical application the procedure for the selection of m will be illustrated.

Selection of c

As in other partitioning clustering methods, in FCR the number of segments, c, is empirically determined by starting with a larger number of clusters and working down to a smaller number. The optimal number of clusters can be determined on the basis of a (scree) plot of the value of the criterion, J_{Rm}, or the percentage of unexplained variance averaged across segments, $1-\bar{R}_a^2$, against the number of clusters. The optimal number of segments is that number where the scree plot, which is similar to that used in factor analysis, levels off, or shows an elbow. In the latter case, a solution with a higher degree of segmentation may have better ability to predict consumer responses within segments.

8.2.4 *Significance testing*

As was noted in 7.2.2, the statistical tests commonly used in regression analysis (F-tests, t-tests) can not be used in clusterwise regression. Because the parameter space increases with the number of observations the asymptotic properties do not apply, and the distribution of F and t values is unknown (Cox and Hinkley 1974). The significance of the regressions within clusters can be examined with simplified Monte Carlo significance tests, developed by Hope in 1968 (see, e.g., Chapter 7, Wedel and Kistemaker 1989, for their use in clusterwise regression). For these tests the null hypothesis is to be rejected if the test criterion from the observed data is greater (or smaller) than at least $M\text{-}M(\alpha/2)$ values of the test criterion calculated from a reference set, consisting of M-1 random samples (α is the level of significance of the two-sided test and both M and $M(\alpha/2)$ are integers). The power of the test increases with M, and approaches that of the uniform most powerful test in the limit (Hope 1968). The reference set for clusterwise regression of preferences can be obtained by randomly permuting preference scores among products and subjects.

8.2.5 *Computer program*

A FORTRAN 77 program, FCRCLUST, was developed to perform the calculations required for FCR analysis. The program requires two input files. One contains the data in free format: subject, product, y variable and x variables. The second input file contains the information required to run FCR: the problem title, the maximum and minimum number of clusters, the value of the fuzzy weight parameter m, the number of parameters in the regression model, the tolerance of J_{Rm} for convergence, the maximum number of iterations allowed, and the seed for the random generator.

FCRCLUST starts from a random (or preset) partition, and performs the iterative estimation process. This process is stopped when the convergence criterion is met, or the maximum number of iterations is reached.
The output includes the regression coefficients, standard errors, t-values and an ANOVA table, for each cluster. The subject memberships can be printed, or saved

for further processing. A plot of the course of J_{Rm} during the iterative process is produced.

The program automatically works down from the maximum number of clusters to the specified minimum number. After the iterative process has been completed for the maximum number of clusters, one cluster is deleted. Each subject's membership value in this cluster is randomly assigned to one of the other clusters. The iterative algorithm is then restarted. This process is repeated until the minimum number of clusters is reached.

An option for Monte Carlo significance testing is included in the program. The number M has to be specified by the user; a frequency distribution of t-values, which can be used for significance testing, is produced.

8.2.6 *A cross-validation procedure*

In order to investigate the validity of FCR on empirical data, we suggest a cross-validation procedure that assesses both the clustering and predictive ability of FCR (see e.g. Punj and Stewart 1983, for a cross-validation procedure in clustering, Berk (1984) for such a procedure used in regression):

1. The data set is split randomly in halves. One half is used as an analysis set, the other half as a validation set.
2. The analysis sample is submitted to FCR (both c and m are assumed to be known). This yields estimates of the coefficients ($\hat{\mathbf{b}}_i^{(a)}$) and memberships ($\hat{u}_{ij}^{(a)}$) in the analysis sample.
3. The y variable in the validation set, $\mathbf{y}^{(v)}$, is predicted from the x variables, using the coefficients of each of the analysis sample segments in turn:

$$\hat{\mathbf{y}}_i^{(av)} = \mathbf{X}^{(v)} \mathbf{b}_i^{(a)}, \tag{8.9}$$

where

$\hat{\mathbf{y}}_i^{(av)}$ = the vector of validation predictions in segment i,
$\mathbf{X}^{(v)}$ = the matrix of independent variables in the validation sample.

4. The distances of observations and predictions are used to calculate the memberships ($\hat{u}_{ij}^{(av)}$) that assign subjects in the validation sample to the analysis sample's segments. The $\hat{u}_{ij}^{(av)}$ are calculated according to Eq. 8.5, with $\hat{D}_{ij}^{(av)}$:

$$\hat{D}_{ij}^{(av)} = \Sigma_k (y_{jk}^{(v)} - \hat{y}_{ijk}^{(av)}) \tag{8.10}$$

5. The validation $R_a^{2(av)}$, the percentage of variance of the validation sample explained by the analysis model, is calculated (Note that $R_a^{2(av)} = R^{2(av)}$ as no adjustment for the number of estimated parameters is necessary.)
6. The $\hat{u}_{ij}^{(av)}$ are used as a starting partition of an FCR analysis of the validation sample. This yields estimates of coefficients, $\hat{b}_i^{(v)}$, memberships, $\hat{u}_{ij}^{(v)}$, and the percentage of variance explained, $R_a^{2(v)}$.
7. Two validation statistics are calculated for each segment, assessing respectively the predictive and cluster validity:
   $S_1 = R_a^{2(av)}/R_a^{2(v)}$. This statistic indicates the cross-validity of FCR predictions.
   $S_2 = r(\hat{u}_{ij}^{(v)}, \hat{u}_{ij}^{(av)})$, which indicates the cross-validity of the cluster solution (r denotes the Pearson correlation coefficient).
8. If the FCR solution obtained in step 6 is a local optimum, the S_1 and S_2 statistics may overestimate the predictive and cluster validity respectively. To investigate the dependence of the FCR analysis of the validation sample on the $\hat{u}_{ij}^{(av)}$ used as a starting partition, in relation to convergence to local optima, the validation sample is analyzed 10 times with FCR using random starting memberships.

8.2.7 *Related methods*

We will now briefly review methods that have recently appeared in the literature, concentrating on one or more of the drawbacks of the "traditional" two-stage procedure (see also Chapter 6).

Within the context of conjoint analysis, Hagerty (1985) proposed a method to overcome the low number of degrees of freedom for estimation at the individual level by combining information across subjects. He claims optimal predictive accuracy for his method. Hagerty estimates the matrix of part-worths by

$\mathbb{C} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Z}$, where $\mathbb{X}$ is the K×P conjoint experimental design matrix, and $\mathbb{Z}$ is the P×n matrix of transformed responses $\mathbb{YSS}'$. $\mathbb{Y}$ is the P×n matrix of responses and $\mathbb{S}$ is a n×c weighting matrix, obtained by factoring the correlation matrix of responses between subjects. $\mathbb{C}$ is the K×n matrix of part-worths, which has rank c. In applications where the $\mathbb{X}$ matrix is the same for all individuals the FCR models show similarity with Hagerty's optimal weighting. It can easily be shown that in FCR the K×c matrix of part-worths within clusters $\mathbb{C}$ is estimated by $\mathbb{C} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Z}$, where $\mathbb{Z}$ is a P×c matrix of weighted totals $\mathbb{YU}^{m}$, where $\mathbb{U}^{m}$ is the n×c matrix of subject memberships u_{ij}^{m}. Hagerty imposes the constraint $\mathbb{S}'\mathbb{S} = \mathbb{I}$, which excludes fuzzy or overlapping clusters. In FCR the constraint $\mathbb{U}'\mathbb{1}_c = \mathbb{1}_n$ is imposed ($\mathbb{1}_c$ and $\mathbb{1}_n$ denote the c×1 and n×1 unit vectors respectively). In optimal weighting, $\mathbb{S}$ is estimated by minimizing the expected mean squared error of prediction, in FCR $\mathbb{U}$ is estimated by minimizing the weighted mean squared error in the estimation sample. FCR is a method developed for segmentation, and the interpretation of the resulting fuzzy benefit segments is more appealing than the interpretation of the factor space resulting from optimal weighting. The interpretation of this factor space is to be aided by e.g. the idealized respondents or the directional cosines method. Although FCR retains the idea that clusters exist, the 'uncertainty' with which respondents are grouped can be influenced by varying m. Both methods suffer from the problem that the dimension of the weighting matrix is to be determined empirically. In 8.4, FCR is compared empirically to Hagerty's method.

Ogawa (1987) and Kamakura (1988) presented hierarchical methods for segmentation in conjoint analysis, which deal with the problem in the two-stage procedures of not maximizing the predictive accuracy. The two methods differ, among other things, in that Kamakura uses least-squares estimation and Ogawa logit estimation. Both start from single-subject clusters which are combined iteratively, to maximize the predictive accuracy, (Kamakura) or to give a minimum reduction of aggregate log-likelihood (Ogawa). As the methods are hierarchical, misclassifications at earlier stages of the algorithms may carry on to higher aggregation levels. Models that are overparameterized at the individual level can not be fitted. For the type of applications discussed in these papers, where the matrix of x variables is the same for all subjects (such as is typically the case in conjoint analyses or for aggregate MDS or factor analysis), overparameterized models ($P > K$) would lead to multicollinearity in FCR analyses, as the rank of both the individual and the pooled **X** matrix is at most equal to K. Thus, if $P > K$, both FCR and the methods of Ogawa (1987) and Kamakura (1988), but also more common techniques such as pooled regression analysis, can not be applied. However, for $K > P$, FCR is less burdened with instability of the estimates of preference parameters since, as compared to estimation at the individual level, it is less sensitive to the measurement error because parameters are estimated within clusters across subjects. (Note that the situation of $P > K$ is not likely to occur often. In aggregate MDS or factor analysis the situation $P > K$ is impossible, while it is not advisable in conjoint analysis.)

The FCR algorithm described in this paper permits consumers to possess partial membership in a number of segments. As such it extends, on a conceptual level[4)], the clusterwise regression algorithms of DeSarbo et al. (1989) and Wedel and Kistemaker (1989, see Chapter 7), which are related to Späth's (1979, 1982) method for clusterwise linear regression. The procedure of DeSarbo et al. includes overlapping clusters, while the algorithm described in Chapter 7 (Wedel and Kistemaker 1989), just like the present method, deals with unequal sets of products for different consumers. An empirical comparison of FCR with the procedure of DeSarbo et al. (1989) is given in 8.4, while a theoretical comparison is presented below.

The clusterwise linear regression methodology described by DeSarbo and Cron (1988) is closely related to the present procedure (see also 4.2.1.2). In their method, estimates of the regression coefficients within a prespecified number of clusters, and of cluster membership parameters, are derived from the assumption of the dependent variable arising from a mixture of conditional normal densities of underlying clusters. The estimates of the model parameters are determined by maximizing the likelihood function, while the membership parameters are estimated via posterior probabilities, using Bayes' rule. The expressions found for $\mathbf{b}_i$ and u_{ij} are similar in form to those in the present paper, whereas the E-M algorithm used by DeSarbo and Cron is related to the Picard iteration. The derivation of the estimates by maximizing the likelihood is methodologically more elegant than the maximization of the criterion J_{Rm} (which is in fact a weighted sum of log-likelihoods), but the asymptotic properties do not apply, as the number of parameter estimates is close to the number of observations, even in the limit (Cox and Hinkley 1974, DeSarbo and Cron 1988). In our methodology, Monte Carlo significance testing is used to overcome this problem. In FCR there is the possibility of choosing different values for the fuzzy weight parameters m, which makes it more flexible with regard to the degree of partitioning of the clusters to be achieved. Moreover, FCR accommodates the analysis of three-way data, whereas DeSarbo and Cron (1988) deal with two-way data, which precludes application in most preference analyses.

Basford and McLachlan (1985) described a mixture method of clustering three-way data, in which one of the modes is clustered on the basis of the other two. However, their method does not allow of simultaneous estimation of regression models within segments.

The fuzzy c varieties family of clustering algorithms (Bezdek et al. 1981a, b) estimate linear principal component models in a prespecified number of clusters, as well as the degree of membership of observations in these clusters (Gunderson 1982). The disjoint principal component models can be used to calibrate a linear regression model within clusters, if both the y and the x variables are included in the data[5)]. The prediction method used is related to latent root regression (Draper and Smith 1976), and does not deal with three-way data.

8.3 Monte Carlo analysis of performance

Before applying FCR to empirical data (8.4 and 8.5), we assessed the performance of the algorithm, with respect to a number of statistics, in a Monte Carlo simulation study. Attention was paid to the following measures of performance:

1. the CPU time required;
2. the number of iterations;
3. the root mean squared error between actual and estimated values of $\mathbf{b}_i$: RMSE(b) ;
4. the root mean squared error between actual and estimated cluster membership: RMSE(u);

5. the difference between the values of J_{Rm} calculated from the actual and those calculated from the estimated cluster memberships: ΔJ_{Rm} (J_{Rm} is in fact the residual sum of squares of the weighted regressions, summed across clusters).

The synthetic data sets were generated with hard partitions (nonoverlapping clusters) of one mode (referred to as subjects), while the addition of random error yielded fuzzy data. The number of clusters (2, 3, or 4), the number of x variables (2, 3, or 5), the number of subjects per cluster (6, 12, or 30), and the number of products (1, 2, or 6) were varied. The y variable was calculated from x variables and coefficients, both randomly generated from a uniform distribution for each cluster, while error, randomly generated from a normal distribution, was added (variances of 0, 25 or 50% of the within-cluster variance of y). These 5 factors were varied according to a 3^5 fractional factorial design in 81 trials. To investigate the performance of the FCR algorithm for varying m, each trial was analyzed with $m = 2$, and $m = 1.5$. Design and dependent measures are based on DeSarbo and Cron (1988). The five dependent measures were analyzed by ANOVA for main effects of the five factors and m, as well as the first-order interactions of m with the other factors, to determine the effect of m on the algorithm performance in different conditions. Table 8.1 shows the results.

In 4 of the 162 trials the algorithm did not converge within the prespecified limit of 100 iterations. On the average, the algorithm required 11 iterations and 96 CPU seconds. From Table 8.1 it appears that a lower value of m increases algorithm performance with respect to all five measures, but the differences are not significant. RMSE(u) was significantly smaller for $m = 1.5$ than for $m = 2.0$ only if no error was present. This result is consistent with the finding of Thrane and Gunderson (1986) that, as m approaches 1, a hard partition is obtained, so that for $m = 1.5$ the actual (hard) partition and the estimated partition approach each other.

In 81.5% of the trials the value of ΔJ_{Rm} was positive, indicating a better reproduction of the data by FCR as compared to the actual clusters.

Table 8.1
Results of the simulation study on FCR performance

Factor	Level	RMSE(b)[1]	RMSE(u)	ΔJ_{Rm}	Iterations[1] (n)	CPU[1] time (s)
Fuzzy weight m						
	2.0	0.16	0.38	110	10.2	80.6
	1.5	0.14	0.34	169	9.6	76.0
x Variables (n)						
	1	0.11^{a} [2]	0.35	58	8.1	72.2^{a}
	2	0.13^{a}	0.34	119	9.7	74.4^{a}
	4	0.22^{b}	0.40	278	10.8	89.3^{b}
Segments (n)						
	2	0.12^{a}	0.36	22^{a}	8.1^{a}	47.4^{a}
	3	0.14^{a}	0.35	119^{ab}	10.7^{b}	85.0^{b}
	4	0.20^{b}	0.38	278^{b}	11.2^{b}	119.0^{c}
Products (n)						
	1	0.38^{a}	0.48^{a}	80	11.6^{a}	73.7
	2	0.13^{b}	0.37^{b}	92	10.8^{a}	78.7
	6	0.04^{c}	0.24^{c}	247	7.6^{b}	82.6
Subjects per segment (n)						
	6	0.24^{a}	0.41^{a}	43^{a}	9.3	59.8^{a}
	12	0.12^{b}	0.36^{ab}	87^{a}	10.1	74.9^{b}
	30	0.11^{b}	0.32^{b}	289^{b}	10.4	107.0^{c}
Error (%)						
	0	0.07^{a}	0.29^{a}	-53^{a}	12.9^{a}	99.9^{a}
	25	0.18^{b}	0.38^{b}	67^{a}	8.4^{b}	70.0^{b}
	50	0.24^{b}	0.42^{b}	405^{b}	8.9^{b}	68.6^{b}

[1] Geometrical averages, due to log transformation before ANOVA.
[2] Factor level averages sharing a superscript are not significantly different at $p < 0.05$.

Moreover, ΔJ_{Rm} increased significantly with the number of segments, the number of subjects per segment, and the percentage of error added. The average value for 0%

error was negative, indicating that the algorithm recovered the actual hard partition only in part of the trials (51.8%). This failure was probably caused by convergence to local optima, related to the tie-breaking rule used.

Recovery of **b** and **U** was better for larger numbers of products and subjects, and a smaller percentage of error. Besides, recovery of **b** decreased with increasing number of parameters. Apparently, as the number of parameters to be estimated increased, problems with multiple optima increased.

The number of iterations required increased with greater numbers of segments, and smaller numbers of products. It was higher when no error was added than for 25% or 50% of error, which might be related to the tie-breaking rule that was entailed in this situation. The same applied to the CPU time used by the algorithm. Further, CPU time increased with increasing numbers of x variables, numbers of segments, and numbers of products.

The conclusions from the Monte Carlo simulation study confirm earlier findings of DeSarbo and Cron (1988) with regard to clusterwise regression performance: as the number of parameters estimated increases, computational time increases and parameter recovery decreases. Increasing the size of the sample (both subjects and products) improves parameter recovery. A higher measurement error decreases parameter recovery, but, presumably because FCR picks up part of the random variation, data reproduction is better than with the actual model and segments. When no error is present, data reproduction decreases and computational effort increases, probably due to problems with multiple optima related to the tie-breaking rule used.

8.4 Empirical comparisons

8.4.1 *Empirical comparison with clusterwise regression*

FCR analysis was compared with the clusterwise regression procedure accommodating overlapping clusters (OCR), developed by DeSarbo et al. (1989), using their data on 'consumer satisfaction'. The aim of the study was to quantify the impact of determinants of consumer satisfaction with respect to eight stock market scenarios, generated from the levels of the five determinants according to a fractional factorial design. The data set comprised measurements of satisfaction among 30 consumers with respect to the eight stock market scenarios presented in a paired comparison format. The hypothesized determinants of consumer satisfaction were: whether the outcome of the purchase of the stock was attributed to oneself (X_1); whether the expectations about the performance of the stock were high or low (X_2); whether, relative to the situation where the stock matches the expectations, the performance of the stock exceeded (X_3) or fell short of expectations (X_4); whether the performance of the stock was high or low (X_5); and whether the inequity of the investor's commission compared favorably or unfavorably to the broker's commission (X_6). For more details about design and data collection we refer to DeSarbo et al. (1989).

The FCR analysis was performed 10 times with different random initial partitions, with $m=1.5$ and $c=2$ (the value of m was chosen by inspecting the within-cluster membership variance for a range of values of m between 1.0 and 2.0). The same solution was found all times within 23 iterations. The solutions of FCR and OCR are shown in Table 8.2.

The correlation between the membership values of FCR and OCR is 0.30 and 0.87 for clusters 1 and 2 respectively. All thirteen subjects with membership in one OCR cluster only had memberships greater than 0.8 in the same cluster in the FCR solution. However, only two subjects (numbers 22 and 30) show substantial membership in both clusters in the FCR solution, whereas seventeen subjects belonged to both clusters according to the OCR results. FCR yields more information on degree of membership than does OCR.

The regression equations found for cluster 1 differ with respect to high performance (X_5), which is more important in the FCR solution. The coefficients for cluster 2 show substantial differences with respect to X_1, X_2 and X_4. The differences between clusters 1 and 2 in coefficients found with FCR are smaller than those found with OCR.

The results of the Monte Carlo tests ($\alpha=0.05$, $M=80$) for testing the significance of the regression coefficients are shown in Table 8.2.

Table 8.2
Comparison of FCR with OCR, consumer satisfaction data

	Cluster membership[1]					Parameter estimates			
	Cluster I		Cluster II			Cluster I		Cluster II	
Subject	OCR	FCR	OCR	FCR	Term	OCR	FCR	OCR	FCR
1	1	0.24	1	0.76	Int	-3.95	-4.34[b]	-0.21	-3.73[b]
2	1	0.19	1	0.81	X_1	5.18[a]	4.56[b]	-3.53[a]	1.09[b]
3	1	0.25	1	0.75	X_2	1.36[a]	1.76[b]	-0.01	1.05[b]
4	0	0.07	1	0.93	X_3	1.03[a]	1.07[b]	-0.60	0.25
5	1	0.93	0	0.07	X_4	-3.45[b]	-3.88[b]	-0.49	-3.31[b]
6	1	0.92	0	0.08	X_5	0.47	1.51[b]	5.10[a]	5.89[b]
7	1	0.95	0	0.05	X_6	1.58[a]	1.73[b]	-0.31	0.84
8	1	0.93	0	0.07					
9	1	0.95	0	0.05					
10	1	0.06	1	0.93					
11	1	0.30	1	0.70					
12	1	0.86	0	0.14					
13	1	0.89	0	0.11					
14	1	0.11	1	0.89					
15	1	0.24	1	0.76					
16	1	0.35	1	0.65					
17	1	0.93	0	0.07					
18	1	0.16	1	0.84					
19	1	0.02	1	0.98					
20	1	0.87	0	0.13					
21	1	0.06	1	0.94					
22	1	0.49	1	0.51					
23	1	0.03	1	0.97					
24	1	0.03	1	0.97					
25	1	0.84	0	0.16					
26	1	0.93	0	0.07					
27	1	0.75	1	0.25					
28	0	0.05	1	0.95					
29	1	0.18	1	0.82					
30	1	0.57	1	0.43					

[1] In OCR, 1 denotes that the subject belongs to the cluster in question, and 0 denotes no membership. In FCR, the figures for the cluster membership reflect the degree to which the subject belongs to the cluster in question.

[a] $p < 0.05$ using the asymptotic standard errors from the inverse of the Hessian matrix (DeSarbo et al. 1989).

[b] $p < 0.05$ using the Monte Carlo test procedure.

All coefficients found with FCR in cluster 1 are significant, whereas in cluster 2 only positive disconfirmation (X_3) and favorable inequity (X_6) are not significantly related to consumer satisfaction.

The FCR analysis confirms the results of DeSarbo et al. with respect to the first segment. In this segment, next to the influences of attribution (X_1), expectation (X_2), disconfirmation (X_3, X_4), and inequity (X_6), which were also established with OCR, the coefficient of (high) performance (X_5) was significant according to the Monte Carlo test procedure in the FCR solution.

The results of the two methods with respect to the second segment showed some differences. Next to performance (X_5), which was also seen to be significantly related to satisfaction in the OCR solution, expectation (X_2) and negative disconfirmation (X_4) were seen to influence satisfaction significantly in the FCR solution. With respect to attribution (X_1), FCR found a positive and OCR a negative coefficient. It may be observed that in FCR the between-segment correlation of estimated coefficients is higher than in OCR, due to the higher degree of overlap allowed in FCR.

It is concluded that the solutions of the methods are consistent and do not yield substantially contradictory conclusions. Unfortunately, information on the predictive fit of OCR was lacking, so that the methods could not be compared with respect to their predictive abilities.

8.4.2 *Empirical comparison with optimal weighting*

Although prediction is not the major issue in FCR, it is worthwhile to investigate the predictive validity of the method. To do this, we compared FCR empirically to the optimal weighting (OW) procedure, proposed by Hagerty (1985), as he claims optimal predictive accuracy for this method. Two synthetic data sets were generated, inspired by the Monte Carlo investigations of Hagerty, for the cases of well defined and diffuse clusters.

Well defined clusters

Three clusters were distinguished with the true part-worths with respect to five three-level attributes as described by Hagerty (1985, Table 1). Part-worths of 40 respondents per cluster were generated by adding normally distributed random numbers (see Hagerty 1985). The responses were generated using the 3^5 design of Addelman's plan 3 (1962) in 16 units. Each attribute was coded as two dummy variables representing the high and low level respectively. These dummies were multiplied by the generated part-worths for each respondent, and a random normal variable was added (see Hagerty 1985). Eight validation trials were constructed for each respondent, to be used for assessing predictive validity.

OW was applied to the estimation sample. A two-factor solution appeared to be optimal, and was used to weight the responses in estimating the part-worths. The estimated part-worths were subsequently used to predict the responses in the validation trials.

FCR was applied with $m = 1.2$ and $c = 3$ (the value of m was chosen by inspecting the within cluster membership variance for a range of values between 1.0 and 2.0). The three generated clusters were recovered with each subject having a high membership in one cluster only. Predicted values of the validation trials were calculated per subject by averaging the predictions from the part-worths across clusters, weighted by membership.

Cross-validated correlations did not differ significantly between FCR (0.581) and OW (0.570), while the percentage of first choices correctly predicted was identical (40.8%). FCR yielded a significantly smaller mean squared error of prediction than OW (0.235 and 0.250, respectively).

Diffuse clusters

The data set with diffuse clusters was generated in a similar way. The part-worths of 40 respondents were now calculated by adding uniform random variables of different ranges (see Hagerty 1985) to each of the three sets of cluster part-worths. The responses of the 16 estimation and 8 validation trials were simulated as before.

Optimal weighting was applied by factoring the correlation matrix among estimation trials. The plot of eigenvalues leveled off at 6 factors. These 6 factors were used to weight the responses in estimating the part-worths, which were then used to predict responses of the validation trials.

FCR was applied with $m = 1.2$. The plot of J_{Rm} and $1\text{-}\bar{R}_a^2$ appeared to level off at 7 clusters. The estimated part-worths within clusters were used to predict the responses of the validation trials as before.

Both the cross-validated correlation and the percentage of first choices correctly predicted were significantly smaller for FCR than for OW (correlations of 0.401 and 0.458 respectively, percentages of 25.8 and 31.7 respectively). However the mean squared error of prediction did not differ between the two methods (0.871 and 0.874 for OW and FCR, respectively).

The results for the three measures of predictive validity found for OW in this study differ somewhat from those found by Hagerty. This is due to some differences in design between our study and Hagerty's: different validation trials were used, and we did not construct the part-worths to be correlated.

In conclusion, the results do not strongly support the superiority of FCR over OW with respect to predictive validity: the performance of FCR equals that of OW for well separated clusters, but is somewhat less for diffuse clusters. The predictive validity of FCR is satisfactory however. It should be noted that FCR was not tailored specifically to conjoint analysis data, nor developed for predictive purposes. Rather, FCR was developed to assess and interpret benefit segments and is, from this point of view, preferable to OW. Also, in contrast to OW, procedures for significance testing are available. Factor analytical procedures lead to clusters that are not easily identifiable (Kamakura 1988), while the approach for cluster interpretation suggested by Hagerty (1985) results in a loss of predictive accuracy. On the other hand, the factor analytic approach is simpler and requires less computational time: FCR took about 15 minutes of CPU time (on a VAX/780), whereas OW took about 5 minutes on the average. Depending on the data collected, the resources available and the purpose of the analysis, either of the two methods may be preferred.

8.5 An investigation into the cross-validity of FCR

8.5.1 *Data*

Products

This application of FCR involves two meat products, cooked ham and salami. In the study, real product samples were used, which strenghtens the external validity of the results. A major meat company produced the samples that were systematically varied on a number of physical and non physical product aspects, on two levels, according to a fractional factorial master design (see Steenkamp 1989 for details). The samples were subdivided into balanced sets of four by means of a blocking procedure. Each subject evaluated one set of four samples. The price of the samples was manipulated at Dfl. 1.89 or Dfl. 2.69 for ham, and at Dfl. 1.39 or Dfl 2.19 for salami. These price extremes are representative of the market situation in the Netherlands.

Measures

Three ratings of overall perceived quality of each sample were obtained, using a Likert scale and a bipolar scale (twice). Price perceptions were measured on a Likert scale. Purchase intention, which was the overall product evaluation of interest in this study, was measured on a bipolar scale. All scales contained seven positions. Ratings for each of the perceived quality measures and for perceived price and purchase intention were normalized across samples for each subject to reduce response-scale bias (Bass and Wilkie 1973). The (normalized) three-item perceived quality measure was reliable (for ham alpha = 0.85, for salami alpha = 0.89).

Subjects

From the consumer panel of a market research agency, a nationwide sample of 480 subjects was drawn. The subjects were randomly assigned to the ham or salami experiment, under the condition that the household to which the subject belonged used the product at least once a month, in order to insure that the subject had some minimum level of experience with the product. Subjects were interviewed at the central test facility of the market research agency.

The subjects invited to participate in the study were the main purchasers of meat products in the household. All subjects were female. They varied in age from 20 to 67, 44.3% had a paid or unpaid job, and the number of members in the household (including the respondent) ranged from one to seven.

8.5.2 *Results*

FCR was applied to the data for both products. (Note that in the present study, the traditional procedure of regressing purchase intention on perceived quality and price is no viable alternative because the regression coefficients are too unstable. Only a single residual degree of freedom is available at the individual level.)

8.5.2.1 *Selection of m*

To investigate the criterion for the selection of m, in relation to the predictive validity of FCR and the number of clusters, the data sets for cooked ham and saveloy were analyzed, with c = 2 and c = 6, and with six different values of m: 1.1, 1.2, 1.3, 1.5, 1.7, 2.0. For each of the 24 solutions obtained, two criteria were calculated: SE_u (SE of those $u_{ij} > 1/2$ for c = 2, and of those $u_{ij} > 1/6$ for c = 6), and $\bar{R}^2_a$ (the percentage of variance accounted for, averaged across segments).

In Figure 8.1 both of these measures are plotted against m. SE_u is optimal at m = 1.3 for c = 2, and at m = 1.5 for c = 6 (in both data sets). $\bar{R}^2_a$ reaches its optimum at m = 1.5 in all cases.

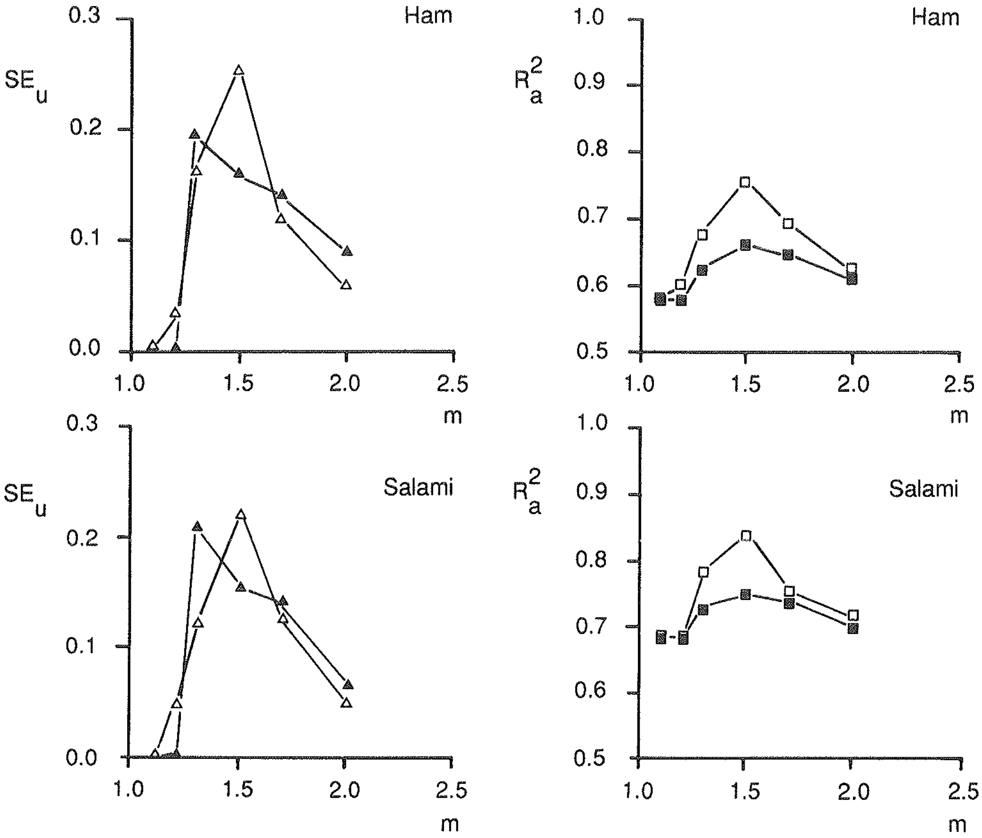

Figure 8.1 Plot of SE_u and R_a^2 against m for the ham and salami data ▲ and ■ indicate the 2-cluster solutions, △ and □ the 6-cluster solutions

It should be noted that the SE_u criterion results in near-optimal values with respect to predictive accuracy. A comparison of the plots shows that both excessive overlap and near-hard partitions, indicated by low values of SE_u, result in a loss of predictive accuracy (Arabie et al. 1981). The value of $m = 1.5$ is used for all further analyses.

8.5.2.2 *Cooked ham*

Estimation of segments

Figure 8.2 plots J_{Rm} and $1-\bar{R}_a^2$ against the number of clusters (two to five). The criteria are expressed relative to the value for the unsegmented solution, i.e., the values of the total sample solution are set at 100%. Lower values indicate a better fit. It appears that the three-cluster solution is the most appropriate.

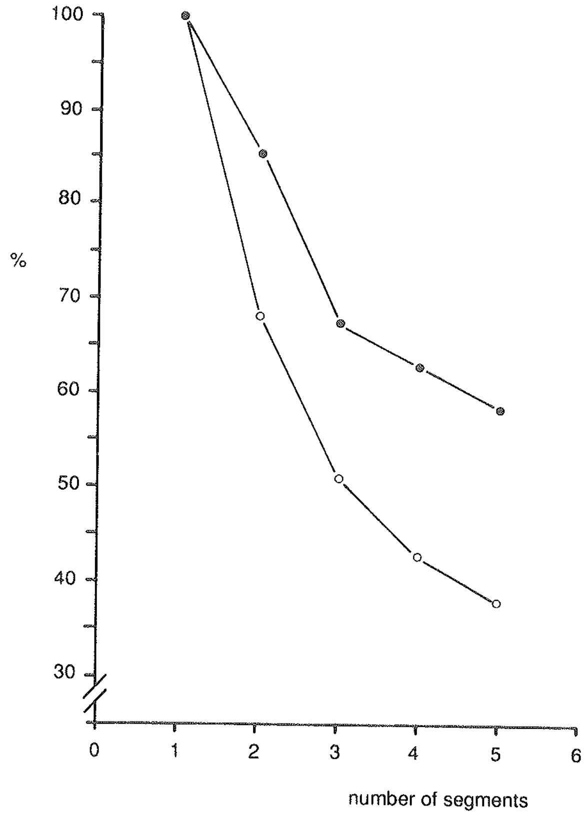

Figure 8.2 Plot of J_{Rm} (o) and $1-\bar{R}_a^2$ (●) (expressed as a percentage of the unsegmented solution) against the number of segments for the ham data

The results of the three-cluster solution for ham are reported in Table 8.3. The coefficients are the unstandardized coefficients of perceived quality and perceived price in the regression on purchase intention, weighted by the segment memberships of individuals. For comparison, the results of the regression analysis for the total sample are reported in the the rightmost column.

Perceived quality is far more important than perceived price in segments 1 and 2, average membership in these two segments being about 0.82. In segment 1 perceived price has a significant, although limited, effect on purchase intention. The effect of perceived quality is smaller in segment 2 than in segment 1, and the price of ham, at least for the prices specified, has no strong effect on purchase intention in this segment. In segment 3 (with an average membership of 0.18) price is more important than perceived quality.

Table 8.3
Results of the three-segment solution of the FCR analysis of data on buying intentions for ham (m = 1.5)

Parameters	Segment 1	Segment 2	Segment 3	Total sample
Constant	-0.147	-0.015	0.259	-0.025
(-10.2, 20.3)[1]	(-14.9)[2]	(-2.0)	(19.1)	(-2.2)
Perceived quality	0.561[a]	0.354[a]	0.242[a]	0.416[b]
(-6.4, 4.8)	(51.1)	(44.7)	(16.6)	(34.2)
Perceived price	-0.095[a]	-0.003	-0.761[a]	-0.149[b]
(-5.4, 4.9)	(-6.0)	(-0.2)	(-36.4)	(-8.1)
R_a^2	0.764	0.687	0.716	0.582
Average membership	0.346	0.471	0.183	1.000

[1] 2.5 and 97.5 percentiles of the distribution of t in the reference set in parenthesis.
[2] t-values.
[a] $p < 0.05$ by the Monte Carlo test procedure.
[b] $p < 0.05$ by the ordinary t-test.

The effect of both product benefits is significant as tested by the Monte Carlo procedure.

Cross-validation

The ham data set was randomly split into an analysis sample and a validation sample. Cross-validation was performed according to the procedure outlined in the previous section; the values of m = 1.5 and c = 3 were used for the analysis. Table 8.4 shows the cross-validation results.

Table 8.4

Results of the FCR cross-validation study for ham (m = 1.5): analysis (AN) and validation (VA) sample coefficients, and validation statistics

	Segment 1		Segment 2		Segment 3		Total sample
	AN	VA	AN	VA	AN	VA	
Constant	-0.019	0.014	-0.166	-0.116	0.192	0.226	-0.018
Perceived quality	0.386	0.328	0.504	0.514	0.330	0.255	0.405
Perceived price	-0.083	-0.042	0.153	-0.079	-0.758	-0.669	-0.146
R_a^2	0.687[a]	0.595[b]	0.626[a]	0.785[b]	0.548[a]	0.660[b]	0.584
S_1		1.153		0.796		0.831	
S_2		0.658		0.576		0.725	

[a] percentage of variance explained in the validation sample by the analysis sample model.
[b] percentage of variance explained in the validation sample by the validation sample model.

The coefficients of the FCR analysis of the analysis sample ($\hat{\mathbf{b}}_i^a$) and the validation sample ($\hat{\mathbf{b}}_i^v$) are fairly similar, although some differences may be observed, especially with respect to the importances of quality in segment 3 and of price in segment 2. (For comparison, the results of the overall regression of the validation sample are given as well.)

The average validation statistic for clustering $S_2 = 0.65$. For the ham data, the clustering validity of FCR is thus satisfactory.

Table 8.4 also shows the percentages of variance explained by cross-prediction and FCR analysis. The $\bar{R}_a^{2(av)} = 62.0\%$. The FCR analysis of the validation sample resulted in $\bar{R}_a^{2(av)} = 68.0\%$. The overall regression of buying intention on price and quality in the validation sample explained 58.8%.

The average of the validation statistic for prediction $S_1 = 0.93$. The predictive validity of FCR for the ham data is thus quite satisfactory; the analysis sample FCR model predicts the responses in the validation sample better than the overall regression model, estimated on the validation sample itself.

From 10 runs of FCR on the validation sample with random starting partitions, the value of $\bar{R}_a^{2(av)}$ reported in Table 8.4 was not surpassed. The solution in Table 8.4 thus does not appear to be a local optimum, and the validation statistics do not overestimate the predictive and clustering validity of FCR.

8.5.2.3 *Salami*

Estimation of segments

The plot of J_{Rm} and $1-\bar{R}_a^2$ against the number of segments for salami is shown in Figure 8.3. A two-segment solution appears to be optimal. The three-cluster solution was also explored but did not yield insights additional to the two-cluster solution.

The parameter estimates of the two-cluster solution for salami are reported in Table 8.5.

Average membership in the segments is 0.40 and 0.60, respectively. In segment 1 both perceived quality and perceived price is more important than in segment 2. However, relative to price, perceived quality is much more important in segment 2, and the effect of perceived price on intention to buy salami is not significant by the Monte Carlo test procedure in this segment. In both segments, around 75% of the variance in intentions to buy is explained. The two segments identified are comparable to the first two segments found for ham, but no segment was found that attached about equal importance to perceived quality and perceived price, such as segment 3 in the solution for ham (where it was also the smallest segment).

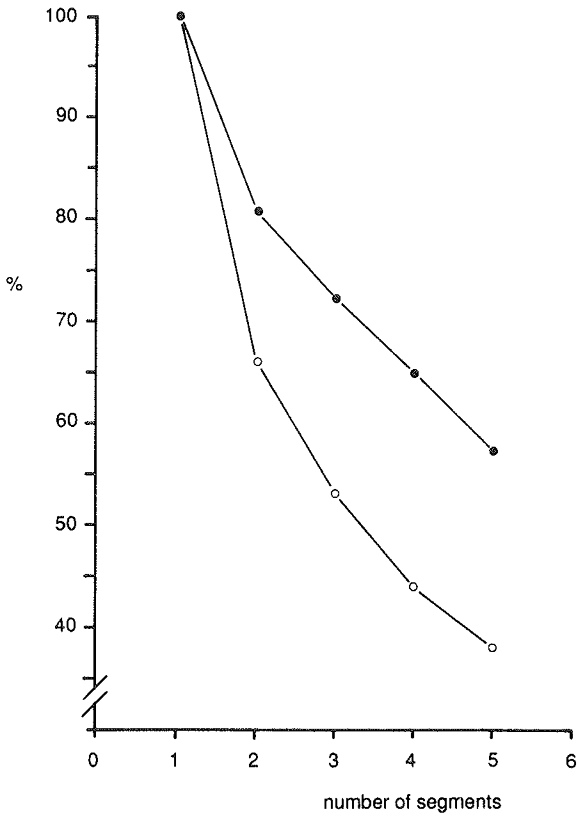

Figure 8.3 Plot of J_{Rm} (o) and $1-\bar{R}_a^2$ (●) (expressed as a percentage of the unsegmented solution) against the number of segments for the salami data

Table 8.5

Results of the two-segment solution of the FCR analysis of data on buying intentions for salami (m = 1.5)

Parameters	Segment 1	Segment 2	Total sample
Constant	-0.186	-0.050	-0.083
(-10.1, 18.7)[1]	(-16.7)[2]	(-6.7)	(-8.7)
Perceived quality	0.647[a]	0.400[a]	0.467[b]
(-3.8, 3.8)	(50.9)	(51.9)	(45.4)
Perceived price	-0.199[a]	-0.003	-0.069[b]
(-4.3, 4.0)	(-10.4)	(-0.0)	(-3.8)
R_a^2	0.755	0.748	0.686
Average membership	0.403	0.597	1.000

[1] 2.5 and 97.5 percentiles of the distribution of t in the reference set in parenthesis.
[2] t-values.
[a] $p < 0.05$ by the Monte Carlo test procedure.
[b] $p < 0.05$ by the ordinary t-test.

Cross-validation

The salami data were randomly split into an analysis and a validation sample. Cross-validation was performed with m = 1.5 and c = 2. Table 8.6 shows the cross-validation results. Again the regression coefficients of the analysis sample and the validation sample are similar.

The average value of the validation statistic for clustering $S_2 = 0.89$, indicating the clustering validity of FCR to be good for the salami data.

The validation $\bar{R}_a^{2(av)} = 69.1\%$. FCR analysis of the validation sample resulted in $\bar{R}_a^{2(v)} = 72.5\%$, the overall regression explaining 64.9%.

The average value of the validation statistic for prediction $S_1 = 0.95$, which indicates that the predictive validity is quite good for this data set.

Table 8.6

Results of the FCR cross-validation study for salami (m = 1.5): analysis (AN) and the validation (VA) sample coefficients, and validation statistics

	Segment 1		Segment 2		Total sample
	AN	VA	AN	VA	
Constant	-0.079	0.009	-0.122	-0.113	-0.071
Perceived quality	0.379	0.450	0.469	0.436	0.446
Perceived price	-0.321	-0.384	0.083	0.145	-0.053
R_a^2	0.669[a]	0.704[b]	0.712[a]	0.746[b]	0.649
S_1		0.951		0.954	
S_2		0.889		0.889	

[a] percentage of variance explained in the validation sample by the analysis sample model.
[b] percentage of variance explained in the validation sample by the validation sample model.

The validation sample was analyzed 10 times with FCR, using random starting partitions. $\bar{R}_a^{2(av)}$ reported in Table 8.6 was not surpassed. The statistics S_1 and S_2 thus do not overestimate the validity of FCR, which had been the case if the solution reported in Table 8.6 had been a local optimum.

8.6 Applications

To demonstrate the practical use of fuzzy clusterwise regression in various fields of segmentation, two applications will be presented. The first application (section 8.6.1) entails an analysis of data on preferences for meat products. In the second application (section 8.6.2) FCR is used to analyze data on consumer attitudes for outlets selling meat.

8.6.1 *An analysis of preferences for meat products*

8.6.1.1 *Introduction*

Consumer preferences for food products are based upon perceived attributes. Wierenga (1983) categorized the attributes of food products into three classes: sensory, instrumental, and expressive attributes. Sensory aspects of food are taste, texture and flavor. Instrumental attributes relate to the functions of foods, such as nutrients, and additives, but also to user-related aspects such as spreadability and packaging. Expressive attributes refer to symbolic aspects, such as exclusiveness, or distinction.

As the importances consumers attach to these different attributes will depend upon their personal circumstances, preferences will vary across consumers, resulting in varying acceptance rates for different products.

In this section, FCR will be applied to analyze the heterogeneity of consumer preferences for meat products used on bread. It will be shown how the results of FCR can guide firms in the development of strategies to improve the position of its meat products in segments of the population, through product modification or communication strategies. Meat is an important product category in the Netherlands, total retail sales in 1988 exceeding 8.5 billion guilders.

8.6.1.2 *Data*

Data on consumer preferences for twelve meat products used on bread in the Netherlands (Steenkamp 1987) were collected in a nationwide sample of 535 subjects, all of which were the main purchasers of food in the households. Each subject was interviewed at home. Subjects were asked to rank the meat products according to their preference. Subjects rated the products on seventeen (2-point) attribute scales, including attributes such as healthy, expensive, natural, and suited for special occasions. Perceived taste was assessed separately on a 7-point scale, because taste is considered to play an important role in preference formation (Steenkamp 1987). Information was also obtained about urbanization, annual household income, age, socioeconomic status and the psychological variable 'locus of control'. The latter concept refers to the degree to which attribution of causality of behavior is made either to oneself or to external sources (Rotter, Chance, and

Phares 1972). Locus of control was measured on a Dutch version of Rotter's (1966) scale (Andriessen 1972). Socioeconomic status was measured by a standardized procedure developed by the Dutch Council of Market Research Agencies (WBO 1985).

Principal components analysis, followed by varimax rotation, was used to reduce the seventeen attribute scales to four perceptual dimensions, explaining 51.7% of the total variance. The four dimensions were judgmentally labeled: fitness for common use, wholesomeness/naturalness, exclusiveness, and fatness/saltness. (For a detailed exposition on data collection and factor analysis results we refer to Steenkamp (1987).) The four perceptual dimensions and perceived taste were to be related to stated preferences, by linear regression. Prior to the regression analysis, preferences and taste ratings were standardized within respondents. Average taste ratings and scores on the four perceptual dimensions for the twelve products are shown in Table 8.7.

Table 8.7

Average scores for taste (TA), fitness for common use (CO), wholesomeness/naturalness (NA), exclusiveness (EX), and fatness/saltness (FA) of meat products

Meat product	TA	CO	NA	EX	FA
1 luncheon meat	-0.652	1.535	-0.327	-0.306	0.044
2 liver	0.144	-0.050	0.891	-0.514	-0.761
3 ham	0.637	-0.172	0.113	0.678	-0.216
4 roast beef	0.803	-0.471	1.077	0.907	-0.567
5 lean bacon	0.129	-0.745	-0.181	-0.191	0.638
6 liverwurst	-0.360	1.026	-0.506	-0.082	-0.085
7 salted meat	-0.423	-0.798	-0.172	-0.305	0.165
8 fat bacon	-0.411	-0.355	-0.186	-0.664	1.107
9 roasted minced meat	-0.640	0.544	-0.503	-0.402	-0.064
10 pâté	0.181	0.033	-0.896	1.149	-0.087
11 cervelat	-0.167	0.296	-0.495	-0.219	0.219
12 smoke-dried beef	0.751	-0.616	1.087	0.523	-0.085

8.6.1.3 *Results and implications*

To reduce the computations required, a random sample of 187 of the 535 subjects was drawn. To determine the value of m, FCR was applied to a range of values of m of 1.0 - 2.0, for c = 2. The pooled within- segment standard errors of the memberships greater than 0.5 were calculated and plotted against m (Figure 8.4).

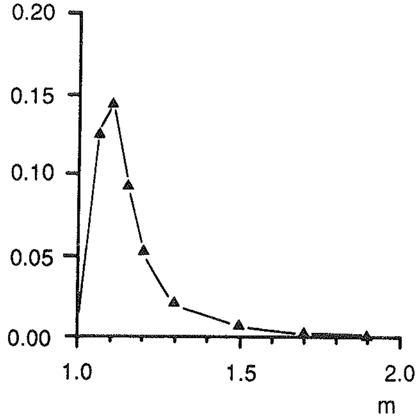

Figure 8.4 Plot of SE_u against m for the meat products data

The plot clearly indicates m = 1.1 to be optimal. (The plot of the within-segment standard error of memberships against m, for c = 3, also indicated this value to be optimal.) The value of m = 1.1 was selected for the following analyses.

FCR analyses were performed, working down from 10 to 2 segments. The plot of J_{Rm} and $1-\bar{R}^2_a$ against the number of segments (Figure 8.5) indicated that a three-segment solution was the most appropriate.

The statistics were expressed relative to the value of the unsegmented solution, i.e. the values of the total sample solution were set to 100%.

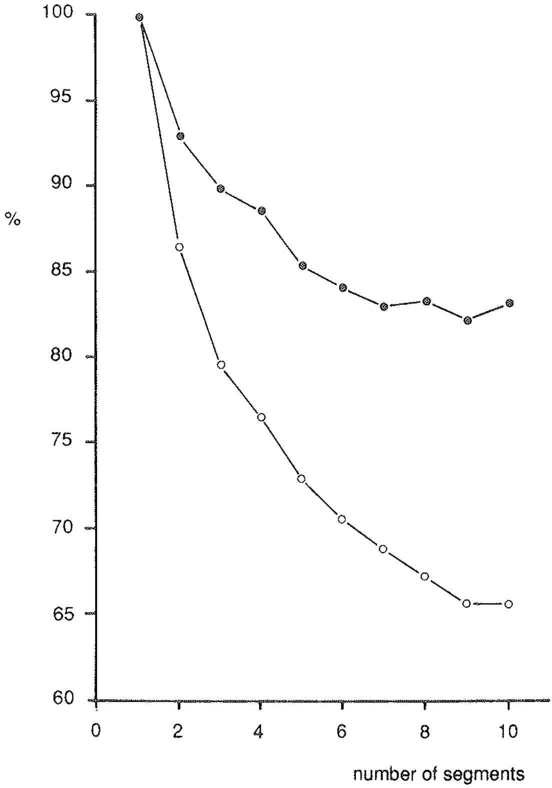

Figure 8.5 Plot of J_{Rm} (o) and $1\text{-}\bar{R}^2_a$ (●) (expressed as a percentage of the unsegmented solution) against the number of segments for the meat products data

The five-segment solution was also explored, but it did not yield insights additional to the three-segment solution, which is also to be preferred on the basis of the plot of the criterion that is minimized in FCR, J_{Rm}. So, the three-segment solution will be presented, all the more since the difference in $\bar{R}^2_a$ is relatively small (56.5 and 54.4 for the five- and three-segment solutions respectively).

Table 8.8 shows the preference functions of the three-segment solution.

Table 8.8

Results of the three-segment solution of the FCR analysis of data on preferences for meat products (m = 1.1)

Product attribute	Segment 1	Segment 2	Segment 3	Total sample
Taste	0.658^{a}	0.703^{a}	0.236^{a}	0.570^{b}
(-4.9; 5.2)[1]	(37.2)[2]	(47.1)	(3.4)	(31.7)
Fitness for common use	0.231^{a}	-0.109^{a}	-0.243^{a}	-0.060^{b}
(-3.1; 4.3)	(15.0)	(-9.3)	(16.4)	(-4.1)
Wholesomeness/naturalness	-0.097^{a}	0.101^{a}	0.274^{a}	0.099^{b}
(-5.0; 2.3)	(-5.9)	(8.3)	(17.6)	(6.4)
Exclusiveness	-0.020	0.079^{a}	0.318^{a}	0.123^{b}
(-3.3; 2.4)	(-1.4)	(6.3)	(21.2)	(8.1)
Fatness/saltness	-0.043	-0.046^{a}	-0.196^{a}	-0.096^{b}
(-3.0; 1.9)	(-2.8)	(-3.9)	(-13.3)	(-6.4)
R_a^2	0.444	0.674	0.514	0.488
Average membership	0.258	0.435	0.307	1.000

[1] 2.5 and 97.5 percentiles of the distribution of t in the reference set.
[2] t-values are given in parenthesis.
[a] $p < 0.05$ by the Monte Carlo test procedure.
[b] $p < 0.05$ by the ordinary t-test.

Table 8.9 contains the weighted averages of product preferences within clusters (weighted with subject memberships), as well as the coefficients of the dummy variable regressions of the logit-transformed memberships, on the consumer characteristics: income (dummy variable coding: 1 = higher, 0 = lower), urbanization (1 = city/suburb, 0 = rural), age (two dummies were used, with the codings: 1 = under 30 years, 0 = 30 years or over; and 1 = 50 years or over, 0 = under 50 years of age), sex (1 = women, 0 = men), locus of control (two dummies: 1 = external, 0 = otherwise; and 1 = internal, 0 = otherwise), and socioeconomic status (0 = higher, 1 = lower).

Table 8.9

Average preferences and logit membership regressions of the three-segment solution of the FCR analysis of preferences for meat products

	Segment 1	Segment 2	Segment 3	Total sample
Average preferences				
1 luncheon meat	0.135	-0.724	-0.965	-0.575
2 liver	-0.096	0.103	0.156	0.068
3 ham	0.675	0.852	0.692	0.760
4 roast beef	0.035	0.993	1.190	0.805
5 lean bacon	-0.052	0.145	-0.066	0.034
6 liverwurst	0.116	-0.529	-0.548	-0.370
7 salted meat	-0.679	-0.402	-0.050	-0.368
8 fat bacon	-0.419	-0.451	-0.548	-0.471
9 roasted minced meat	-0.132	-0.556	-0.524	-0.439
10 pâté	0.008	-0.011	0.106	0.027
11 cervelat	0.132	-0.353	-0.410	-0.252
12 smoke-dried beef	0.227	0.891	0.998	0.752
Logit membership regression coefficients				
Residence	0.729	-0.757	-0.013	
Sex	-2.330[a]	0.768	0.820	
Income	0.364	0.409	-0.766	
Age < 30	1.532[b]	-0.957	-0.950	
Age > 50	-0.802	-1.452[b]	2.084[b]	
Internal locus of control	-1.243	0.806	-0.582	
External locus of control	0.459	1.819[a]	-3.041[a]	
Socioeconomic status	-0.696	0.663	-0.391	
R^2_a	0.030	0.030	0.066	

[a] $p < 0.05$.
[b] $p < 0.10$.

In segment 1, containing 26% of the sample, taste has the highest importance weight. Further, fitness for common use has a significantly positive and wholesomeness/naturalness has a significantly negative relation to preference (Table 8.8). Preferences for product 3 (ham) are high, and no strong competitors of this product are present. Compared to segments 2 and 3, preferences for products 1, 6, and 11 are relatively high, which indicates opportunities for these products in segment 1 (Table 8.9). As these products rate low on taste, product modification may be used to increase the taste rating for these products (Steenkamp and van Trijp 1989). Whereas products 1 and 6 are perceived fit for common use, product 11 is not (Table 8.7), and the perceived fitness for common use of this product might be enhanced by promotion. As there is relatively little competition within this segment, possibilities for foreign competitors exist, as consumers may look for variety. Men and people under 30 years had higher memberships in this segment (Table 8.9).

Segment 2, the largest segment, contains about 44% of the sample. Taste overridingly determines consumer preference. Further, consumer preference increases with exclusiveness and wholesomeness and decreases with fatness/saltness and fitness for common use (Table 8.8). Products 3, 4, and 12, rating high on taste and wholesomeness, are strong competitors, as indicated by high preferences within this segment. Preferences for product 5 are high as compared to segments 1 and 3, which indicates opportunities (Table 8.9). As this product rates high on fatness/saltness, marketing strategy for this product should aim at decreasing consumers' perceptions of this attribute (Table 8.7). Products 1, 6 to 9, and 11 have low preferences, mainly because of a negative taste appeal. Consumers with a more external locus of control had higher memberships, while consumers over 50 had a lower membership in this segment (Table 8.9).

Table 8.8 shows that in segment 3 all five product dimensions exert an almost equally strong influence on consumer preference. The direction of the effects is the same as in segment 2, but the role of taste is less, and that of the other dimensions more prominent. Less fat and salt, higher perceived wholesomeness, exclusiveness and taste, and lower fitness for common use are associated with higher preference. Products 4 and 12, and to a lesser extent 3, are strong competitors (Table 8.9). Preferences for product 7 are higher than in the other two segments, indicating possible opportunities for this product. Low preferences are found for products 1, 6, 8, 9, and 11, which have low scores for taste, wholesomeness, and exclusiveness (Table 8.7). Product modification or correction of mistaken perceptions of these attributes will increase preferences in this segment. Promotion of aspects related to exclusiveness, naturalness, and wholesomeness, and reduction of fat and salt content, as well as more exclusive packaging, may increase consumer preferences in this segment, in which consumers are older and have a more internal locus of control (Table 8.9).

Although the relationships of segment memberships with consumer characteristics are weak, as indicated by low R_a^2, the consistent relationships are worth noting. The age of consumers increases from segment 1 to 3, while the importance of taste decreases and the importance of health-related aspects increases in that direction. Moreover, the older and more health-oriented consumers in segment 3 were found to attribute the causality of their behavior more to themselves than is the case in the other segments.

This application has shown that FCR results broaden our insight into the importance of attributes of meat products in segments of the market. FCR thus provides information that supports a companies (single or multiple) benefit positioning of their products. The FCR analysis was shown to reveal the competitive structure within segments, and suggested strategies of product modification and communication that were not apparent from the unsegmented solution.

The degree to which the segments can be effectively reached, depends on the extent to which consumers that have a high segment membership can be profiled with variables that indicate where they live, where they shop, and to which media they are exposed. It was shown that the segments revealed by FCR can be made accessible by relating consumer memberships in a segment to demographic, socioeconomic and psychographic variables, in a second step of the analysis. The profiles of consumers that have high memberships in the segments in question can be used in designing communication strategies that appeal to the target group, as well as in the choice of media through which the segments are to be reached.

8.6.2 *An analysis of attitudes' for outlets selling meat*

8.6.2.1 *Introduction*

A critical aspect of retailers' ability to maintain their market position is to develop and maintain a favorable store image. Store image can be defined as an overall attitude towards the store, based upon the perceptions of relevant store attributes (Bearden 1977; Doyle and Fenwick 1974; James et al. 1976). Image considerations are an important aspect in the development of an integrated marketing strategy for individual stores, store chains, and shopping centers.

However, different groups of consumers might place different importances on the various store attributes (Martineau 1958) and, ideally, the image attributes stressed by the store should be those to which the target segment attaches the most importance. The importance of market segmentation on the basis of the store image attributes and the development of an image that conforms to the needs of the store's target group of consumers have been repeatedly stressed in the literature (Doyle and Fenwick 1974; Hansen and Deutscher 1977; James et al. 1976; Verhallen and DeNooy 1982).

In several of studies, the existence of consumer segments, differing in importance attached to various store attributes, was investigated. Gentry and Burns (1977), Hansen and Deutscher (1977), and Schiffman et al. (1977) defined the basis of segmentation a priori, and subsequently explored whether differences in store attribute importances exist between segments.

More recently, some researchers have used the two-stage approach to benefit segmentation. Tantiwong and Wilton (1985) segmented consumers on the basis of directly rated importance of store attributes. In one of the most elaborate segmentation studies published to date, Verhallen and DeNooy (1982) clustered consumers on the basis of idiosyncratic importances, which were estimated with conjoint analysis.

The problems related to the two-stage approach have been summarized above. A problem, however, that is even more pregnant in store image research than in product research is that the number of alternatives in a certain category the consumer is aware of is typically quite small (e.g., Goldman 1977). It has even been argued that consumers should only rate stores they currently patronize (Schiffman et al. 1977), thus limiting further the number of observations.

In the present section, fuzzy clusterwise regression will be applied to study the different bases of store image toward outlets selling meat. It will be demonstrated that fuzzy clusterwise regression yields important insights into the market that can be used for developing a retail strategy.

8.6.2.2 *Data*

In two cities in the Netherlands, 148 consumers were interviewed at their homes using the computer-interactive interviewing program Ci2 (Sawtooth 1986). All subjects were the main purchasers of meat in their households. Data were collected for the store in which the subject bought most meat[6)].

The following five store image attributes were identified: product quality (7 items), pricing (24 items), service quality (6 items), store atmosphere (6 items), and assortment (5 items). The service items dealt with personnel, hygiene, speed of service and checkout time, ease to order, and the way complaints are dealt with. The items operationalizing store atmosphere included the type of people one meets in the store and the store's interior and exterior. Overall store image was measured with seven evaluative items.

In addition, information was obtained about the weekly expenditure on meat and meat products, the number of shops (occasionally) patronized, and the sociodemographic characteristics sex, age, family size, level of education, and employment status. Further, store involvement was measured using a reduced version (Zaichkowsky 1987) of the Personal Involvement Inventory (Zaichkowsky 1985), consisting of 10 bipolar items. Involvement with meat was measured using a modified version of Kapferer and Laurent's (1985) involvement scale, consisting of 14 Likert scale items.

Responses on overall store image, store attribute and involvement items were measured on 50-point graphical scales. The reliabilities of the scales for each of the constructs range from 0.645 to 0.908, which is considered adequate (Nunnally 1967).

8.6.2.3 *Results and implications*

Average ratings on the store image attribute scales were used as input to FCR.

The plot of standard deviations in membership values greater than 0.5 ($c = 2$) against m is shown in Figure 8.6. The curve is rather flat for values of m between 2.0 and 3.0, with the optimum at $m = 2.8$; this value of m was chosen for further analyses.

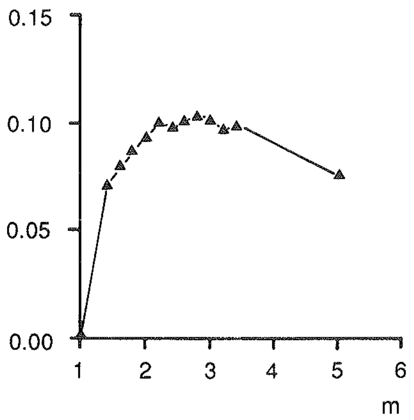

Figure 8.6 Plot of SE_u against m for the store image data

Figure 8.7 shows the plot of J_{Rm} and $1-\bar{R}_a^2$ against the number of segments. It appears that the three-segment solution is the most appropriate as the plots show an elbow at this point. The results of the three-segment solution are reported in Table 8.10.

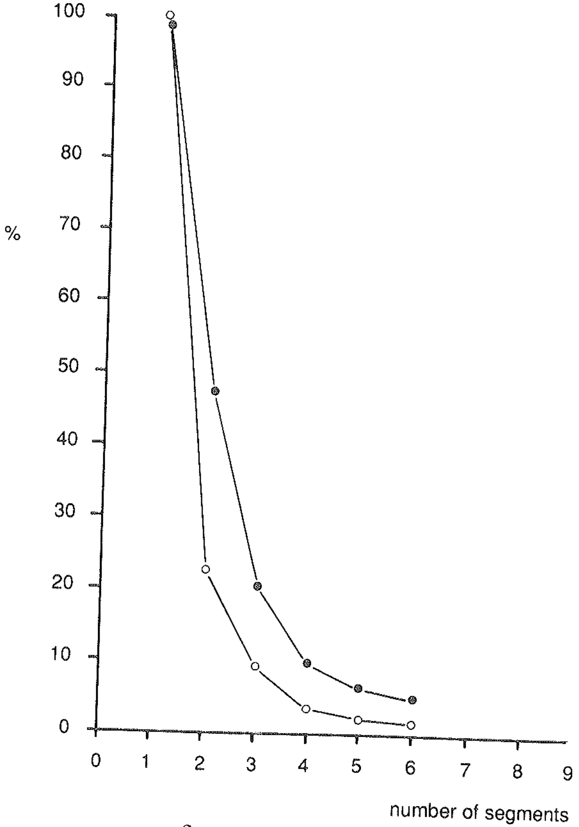

Figure 8.7 Plot of J_{Rm} (o) and $1-\bar{R}_a^2$ (●) (expressed as a percentage of the unsegmented solution) against the number of segments for the store image data

The first segment, with an average membership of 32.2%, largely based store image on the tradeoff between product quality and price. This segment represents shoppers looking for value for money, and was named judgmentally 'value shoppers'. In the largest segment, comprising 40 % of the sample, store image was predominantly based on product quality. None of the other store attributes had a significant influence although a high price (price as quality index) and good service contribute to store image. This segment was named the 'quality shoppers'. The third segment exhibits significant effects for product quality, service, and store atmosphere. The direction of the effect of quality (-2.050) is counterintuitive, but examination of the weighted within-cluster correlation matrix of store attributes revealed a relatively high correlation between quality and service (0.179). This multicollinearity has severely affected the estimate of the regression coefficient of quality (cf. Pedhazur 1982). Simple linear regression weighted with memberships for this segment yielded a moderately large positive coefficient for quality (0.730), while the coefficient for service appears to be relatively unaffected by the correlation with quality, its simple weighted regression coefficient being 1.310. In this segment, judgmentally named 'service shoppers', the effect of assortment on store image is also substantial, although not significant.

Thus, segmenting the market adds substantially to our insight into the importance of different bases of store image, which is important for retail strategy. The total-sample analysis (see Table 8.10) suggests that store image is only based on product quality and service. However, analysis of the segments reveals that price has a significant negative effect on store image for about one third of the subjects. Further, one segment attaches considerable importance to store atmosphere and to a lesser extent to assortment. Thus, segmenting the market suggests additional retail strategies that are not apparent from the unsegmented solution (see also below).

The relationships between segments and consumer characteristics were explored with partial least squares regression (PLS; Wold 1982, Martens and Martens 1986). PLS estimates a small number of latent factors to express the systematic variation in the predictor set (consumer characteristics) that is related to the criterion variables (segment membership). Loadings are calculated for the criterion and predictor variables, indicating their relationship with the latent factors.

The number of PLS factors to be retained is determined by looking for a maximum in the plot of explained variance in the criterion variables, under the condition that the factors are significant. PLS is not burdened with multicollinearity among the predictor or criterion variables.

A number of sociodemographic characteristics were included as dummy variables: level of education (1 = higher, 0 = lower), sex (1 = female, 0 = male), and employment status (1 = has a job outside the home, 0 = has no job outside the home). Membership values were logit-transformed before PLS analysis.

Table 8.10
Results of the three-segment solution of the fuzzy clusterwise regression analysis of store image data (m = 2.8)

Store attribute	Segment 1	Segment 2	Segment 3	Total sample
Product quality	1.222[a]	1.130[a]	-2.050[a]	0.711[b]
(-19.5, 19.7)[1]	(20.3)[2]	(24.9)	(-20.6)	(4.8)
Price	-1.614[a]	0.806	0.363	-0.086
(-21.2, 19.2)	(-21.8)	(11.6)	(4.5)	(-0.5)
Assortment	0.098	0.218	0.467	0.144
(-21.6, 23.7)	(3.1)	(5.9)	(12.3)	(1.7)
Service	-0.103	0.648	1.529[a]	0.638[b]
(-22.0, 171)	(-1.7)	(12.9)	(20.4)	(4.6)
Atmosphere	0.283	-0.090	0.698[a]	0.089
(-31.2, 10.5)	(8.9)	(-2.6)	(14.7)	(1.0)
R_a^2	0.881	0.948	0.902	0.536
Average membership	0.322	0.395	0.283	1.000

[1] 2.5 and 97.5 percentiles of the distribution of t in the Monte Carlo test procedure in parenthesis.
[2] t-value in parenthesis.
[a] $p < 0.05$ by the Monte Carlo test procedure.
[b] $p < 0.05$ by the ordinary t-test.

The percentage of variance explained in cluster memberships showed a maximum at three factors, all of which were significant by cross-validation.

Table 8.11 shows the results. We will concentrate on salient loadings (>0.4) for the consumer characteristics.

Factor 1 indicated that quality shoppers are the most involved in the store where they buy meat, whereas value shoppers are the least involved. Quality shoppers are predominantly female, and mostly have no job. Factor 2 shows that, as compared to quality and service shoppers, value shoppers spend less on meat and meat products and have smaller families. Service shoppers are distinguished from quality and value shoppers on factor 3. Most interestingly, service shoppers tend to be more store-loyal, in that they patronize fewer shops. Whereas other stores may offer quality or value, service is an intangible asset that may be difficult to duplicate. Service shoppers are less often female and, interestingly, tend to be less involved in the store.

Table 8.11

Loadings of segment membership and consumer characteristics on three PLS-factors

Variables	Factor 1	Factor 2	Factor 3
Segment membership			
1 Value shoppers	0.101	0.151	-0.138
2 Quality shoppers	-0.200	0.031	-0.084
3 Service shoppers	-0.062	0.073	0.195
R^2	0.022	0.008	0.013
Consumer characteristics			
Involvement with store	-0.500	0.231	-0.402
Involvement with meat	-0.240	0.153	0.009
Number of shops patronized	0.324	-0.262	-0.674
Amount spent on meat	-0.287	-0.555	-0.319
Amount spent on m. products	-0.298	-0.488	0.013
Level of education	0.192	-0.354	-0.067
Sex	-0.350	0.011	-0.586
Age	-0.227	0.368	0.098
Family size	-0.157	-0.525	0.050
Employment	-0.487	-0.051	0.302
R^2	0.070	0.104	0.026

The three segments distinguished provide meat retailers with major opportunities for developing differentiated appeals. The largest segment consist of quality shoppers. Quality shoppers will be especially receptive to a marketing strategy that includes selling a larger variety of meat cuts considered to be of higher quality, as well as better quality of certain meat cuts. Higher prices need not deter these consumers. Interestingly, quality shoppers tend to spend more on meat than the other two segments. One segment attaches great importance to service, and to a lesser extent to atmosphere. Retailers aiming at this segment should put special emphasis on hiring and training personnel that has a broad knowledge of meat and has the 'right' attitude toward other service aspects such as hygiene and speed of service.

Despite recent trends towards quality and service, a sizeable segment is still interested in value for money. These shoppers could be attracted by a marketing strategy aimed at reasonable quality at low prices. Frequent special offers are a relevant element of this strategy.

8.7 Conclusions

Although the problem is not unique to FCR, a note on the choice of the number of segments seems appropriate. As with other partitioning methods, in FCR the number of clusters is determined empirically, using J_{Rm} or $1\text{-}\bar{R}_a^2$ as heuristic measures. Theoretically it may be of interest how many segments are present in the population. It is questionable, however, whether in practice an exact number of segments can be pinpointed as giving the best representation of the variability among consumers, even if segments are fuzzy. From the point of view of the marketing manager, the number of segments decided upon is dependent on the size of the segments, the marketing budget, the ability to cater different segments and the amount of detail in information required (see Chapter 2). The plots of J_{Rm} and $1\text{-}\bar{R}_a^2$ against the number of segments provide an indication of the increase in information when the number of segments is increased, which may assist the marketer in his trade-off of information and the cost of marketing strategies. Searching for the optimum number of clusters may require a great deal of CPU time for large data sets, especially because for each number of clusters the analyses should be repeated to avoid local optima.

Computational requirements at this stage of the analysis may be reduced by using a random subset of subjects (usually the largest mode of the data), although its effect on the number of clusters is unknown. Alternatively, an efficient initial partition based, for example, on a clustering of preferences could be used.

A second judgmental choice that has to be made in the application of FCR, is the choice of the fuzzy weight parameter m. This parameter provides flexibility with respect to the degree of partitioning, and influences memberships and the differences in estimated regression coefficients between clusters. The within-cluster standard error of memberships larger than 1/c was suggested as an empirical measure to select m, as it guards against nonoverlapping and excessively overlapping solutions, which was demonstrated in the empirical applications. In practice, visual inspection of within-cluster memberships for a few values of m will suffice for the selection of an optimal value. In the applications, differences in predictive fit resulting from different values of m around the 'optimum' were small. It should be noted that the criterion suggested is not optimal from a statistical point of view in that it minimizes J_{Rm} or maximizes the predictive fit.

A cross-validation procedure was suggested to assess the stability of FCR solutions. The procedure addresses both the predictive and clustering ability of FCR, and is easy to use in practice. The cross-validation results of the two data sets on price-quality tradeoff for meat products were promising. The cross-classified memberships showed high correlations with memberships obtained from FCR analyses. The validation predictions reached around 95% of the values of the predictive fit of the FCR solutions, and surpassed the fit of overall regression models fitted on the validation data. There is thus evidence that FCR may display considerable cluster and predictive validity.

The empirical studies presented support the usefulness of FCR as a segmentation technique. The results of the studies have suggested marketing strategies that were not apparent from the unsegmented analyses. A substantial improvement was found in the accuracy in predicting consumers' overall evaluations on the basis of attribute perceptions, and considerable differences in benefit importances between segments were revealed, which could be translated into marketing strategy. Although the percentage of variance explained was small, the (significant) relationships of segment memberships with consumer characteristics were consistent and supported the validity of the solutions. These findings support the viability of FCR as a segmentation technique in relation to multiattribute models of consumer decision-making.

It is important to note that in FCR no assumptions are made on the distribution of the data, since significance tests are based on Monte Carlo test procedures. However, in some instances, for example when choice frequency data are collected, a suitable transformation of the dependent variable, such as a logit transformation, may be advisable in view of the interpretation of the estimated coefficients. Fuzzy clustering could be applied to analysis of variance as well, so that data with nested error structure can be analyzed with mixed models and fuzzy segmentation performed simultaneously.

In conclusion, FCR is a powerful method for benefit segmentation within the framework of preference formation. It provides information on strategic issues concerning market segments, on the degree of competition between products or brands within these segments, and on opportunities presented by gaps in the market, all from the consumer's perspective. The procedure yields identifiable and substantial segments, while providing an understanding of the role of product dimensions in the formation of preferences or intentions. The method overcomes the arbitrariness of defining nonoverlapping product market segments, and can be used for predictive purposes. Variation in preference formation, for instance across different usage situations, is accommodated by allowing consumers to exhibit multiple preference functions. In applications, attention should be paid to proper significance testing using Monte Carlo test procedures, and to the problems of local optima.

9 GENERALIZED FUZZY CLUSTERWISE REGRESSION: A METHOD FOR SIMULTANEOUS FUZZY MARKET STRUCTURING AND BENEFIT SEGMENTATION

9.1 Introduction

Strategic issues in planning firms' marketing efforts are critically dependent upon the definition of the market and its structure. A product market structure is defined to be a group of products for which similar patterns of benefits are sought by a particular group of customers for specific occasions, and which are consequently judged to be substitutes (Day et al. 1979). The definition of market structures is conceptually similar to the identification of a market segment (Lilien and Kotler 1983), and it has been argued that market segmentation and market structuring are complementary (Grover and Srinivasan 1987), focusing on the demand and supply side of a market respectively. The methods that have been used for defining product markets have been classified into purchase or use behavior approaches, and perception or judgmental approaches (Day et al. 1979). In the former approaches (cross)elasticities of demand, similarities in behavior, interpurchase times (Fraser and Bradford 1983, Grover and Rao 1988) and brand switching (Grover and Srinivasan 1987) have been used as a basis for market definition. The first two bases are rarely used.

Interpurchase times and brand switching as potential measures are based upon the assumption of stable switching behavior and are limited to markets with high repeat rates. The perceptual and judgmental approaches entail decision sequence analysis, perceptual mapping and consumer judgment of substitutability. These approaches have seen wide use in market definition studies (Day et al. 1979), as it is generally recommended to define markets on the basis of the view of their customers (Hruschka 1986). Cluster analysis has been applied to determine competitive market structures by clustering brands. The analyses often parallel the two-stage procedures in benefit segmentation (Arabie et al. 1981).

Two of the three limitations that affect the validity of the traditional two-stage procedure for benefit segmentation, the use of clustering procedures not maximizing the predictive fit and estimating nonoverlapping clusters, also hold for the two-stage procedure applied to market structuring. The third limitation, unreliability of the estimates of importances, is less relevant in market structuring, as a large number of observations are usually available to estimate the importances for each brand.

With respect to non-overlapping clusters of brands, brands may compete with different subsets of brands (Arabie et al. 1981), and therefore belong to more than one cluster. A number of approaches have been proposed that address the issue of overlapping clusters. The ADCLUS model proposed by Arabie et al. (1981) derives overlapping cluster solutions from similarity data, and was applied to market structuring by Srivastava et al. (1984). Hruschka (1986) suggested the use of fuzzy clustering methods such as fuzzy c-means for market structuring and market segmentation. He demonstrated empirically that these methods provide a more valid cluster solution than the nonoverlapping or the overlapping procedures.

The clusterwise regression procedures proposed by DeSarbo, Oliver and Rangaswamy (1989), DeSarbo and Cron (1988) and Wedel and Steenkamp (1989, see Chapter 8) estimate overlapping and fuzzy clusters and maximize the predictive fit simultaneously. These procedures will be discussed in more detail in section 9.2.5.

There is, however another limitation of current segmentation techniques not discussed thus far, which is that they do not analyze the structure of the market in relation to benefit segments. Brands may compete in different subsets of brands on the basis of different benefits desired by different segments. This means that the competitive market structure depends on consumer segments (Green, Wind and Claycamp 1975).

Hruschka (1986) used the following two-stage approach to fuzzy market structuring and segmentation. First, fuzzy market segments were derived from between-subject similarities. Second, the fuzzy partition was transformed into a hard partition, and a fuzzy classification of products was obtained within each of these 'hard' segments. The methods used by Hruschka (1986) operate on between-subject or product similarities.

Grover and Srinivasan (1987) described a method for simultaneous market structuring and market segmentation that estimates segment sizes and within-segment market shares by a latent class analysis of the cross-classification matrix of numbers of brands purchased on two occasions. The market is segmented into brand-loyal and fuzzy brand-switching segments. The method was generalized to account for non-stationarity in the within-segment market shares over the time horizon considered (Grover and Srinivasan 1989). The methods of Grover and Srinivasan (1987, 1989) are tailored to the analysis of purchase or use behavior, segments being defined as groups of consumers with homogeneous purchase probabilities.

The method recently proposed by Kamakura and Rusell (1989) simultaneously estimates segments and coefficients of price of a logit model within segments, from brand choice data. Their choice model partitions the market into segments differing in both brand preference and price sensitivity. The market structure within segments is described in terms of choice shares that are linked to preferences and price elasticities. This method also falls in the class of approaches for the analysis of purchase behavior.

In this chapter we propose a method that integrally addresses the limitations of the traditional procedures of market structuring and benefit segmentation. The method proposed, generalized fuzzy clusterwise regression (GFCR), relies upon judgmental data, segments being defined as groups of consumers that are homogeneous in preference functions. GFCR is a generalization of FCR in that fuzzy market structuring is incorporated, allowing brands as well as subjects to have memberships in several clusters. In GFCR subjects can belong to more than one segment when they attribute different importances to product dimensions, for example depending on the usage context (Miller and Ginter 1979). Correspondingly, brands may compete with different subsets of brands on different dimensions, depending on segments. GFCR is a generalization of FCR, the method described in Chapter 8. The analysis of brand preferences with GFCR may serve as a basis for strategic marketing planning, as it pictures both the opportunities and threats facing a business. GFCR provides insight into the reasons of current and potential patterns of competition among brands within segments, from the consumers' point of view.

Section 9.2 of this chapter describes the proposed method. Section 9.3 contains a Monte Carlo investigation of the performance of GFCR on synthetic data. Section 9.4 entails a comparison of GFCR with fuzzy clusterwise regression (FCR, Chapter 8) and clusterwise regression (CR, Chapter 7). In section 9.5 an application is given to data on butter and margarine, and the cross-validity of GFCR is investigated. In Section 9.6 managerial and research issues are discussed.

9.2 The method

9.2.1 *Generalized fuzzy clusterwise regression*

The data for the analysis are assumed to consist of preferences of subject j (j=1...n) for product k (k=1...K_j; $1 \le K_j \le K$, where K is the number of brands included in the study, and K_j the number of brands evaluated by subject j). The preferences are to be related to P perceived product dimensions or profile attribute levels. Assume that there exists a fixed number of clusters, c, which is known, and each of which has a unique preference function. Assume further that brands and subjects can be a member of the same set of clusters (as subsets of brands compete within segments). For cluster i (i=1...c, $2 \le c < \min(n,K)$) the model that relates theoretically reconstructed preferences to product dimensions is:

$$\mathbf{h}_i = \mathbf{X}\mathbf{b}_i, \tag{9.1}$$

where

$\mathbf{h}_i$ = the (N×1) partitioned vector of theoretically reconstructed preferences ($N=\Sigma_j K_j$), consisting of the n (K_j×1) subvectors $\mathbf{h}_{ij}$,

$\mathbf{X}$ = the (N×P) matrix of perceived product dimensions or profile attribute levels, accordingly partitioned,

$\mathbf{b}_i$ = the (P×1) vector of importances,

$\mathbf{e}_i$ = $\mathbf{y} - \mathbf{h}_i$ is an (N×1) vector of independent error terms, e_{ijk},

$\mathbf{y}$ = the vector of observed preferences.

Let $\mathbf{U}$ denote a (c×n) matrix with elements u_{ij} ($0 \le u_{ij} \le 1$), representing a fuzzy c-partition of subjects:

$$\Sigma_i u_{ij} = 1; \qquad \Sigma_j u_{ij} > 0, \tag{9.2}$$

and $\mathbf{T}$ a real (c×K) matrix with elements t_{ik} ($0 \le t_{ik} \le 1$), representing a fuzzy c-partition of products:

$$\Sigma_i t_{ik} = 1; \qquad \Sigma_k t_{ik} > 0. \tag{9.3}$$

The purpose is to estimate the c-partitions $\mathbf{U}$ and $\mathbf{T}$, and the c parameter vectors $\mathbf{b}_i$. A weighted sum of squared error criterion J_{Rml} is defined:

$$J_{Rml} = \Sigma_i \Sigma_j \Sigma_k t_{ik}^{l} u_{ij}^{m} e_{i\,jk}^{2}, \tag{9.4}$$

where the summations are across the appropriate values. The parameters in the exponent of t_{ik} and u_{ij}, l ($l > 1$) and m ($m > 1$) respectively, are fixed weights, which to influence the extent to which products (l) or subjects (m) belong to more than one cluster (Thrane and Gunderson 1986).

The estimates of the parameters t_{ik}, u_{ij}, and $\mathbf{b}_i$ are obtained by minimizing J_{Rml}, under the sum constraints (9.2) and (9.3), given c, m, and l (see Appendix B):

$$\hat{\mathbf{b}}_i = (\mathbf{X}'\hat{\mathbf{T}}_i^{l}\hat{\mathbf{U}}_i^{m}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{T}}_i^{l}\hat{\mathbf{U}}_i^{m}\mathbf{y}, \tag{9.5}$$

$$\hat{t}_{ik} = 1/\Sigma_h (\hat{G}_{ik}/\hat{G}_{hk})^{1/(l-1)}, \tag{9.6}$$

$$\hat{u}_{ij} = 1/\Sigma_h (\hat{D}_{ij}/\hat{D}_{hj})^{1/(m-1)}, \tag{9.7}$$

where $h = 1 \ldots c$, and

$$\hat{G}_{ik} = \Sigma_j \hat{u}_{ij}^{m} \hat{e}_{i\,jk}^{2}, \qquad \hat{D}_{ij} = \Sigma_k \hat{t}_{ik}^{l} \hat{e}_{i\,jk}^{2}, \tag{9.8}$$

$\mathbf{T}_i$ and $\mathbf{U}_i$ are partitioned diagonal matrices:

$$\hat{\mathbf{T}}^{l}_{i} = \mathrm{Diag}(\hat{\mathbf{T}}^{l}_{ij}), \qquad \hat{\mathbf{T}}^{l}_{ij} = \mathrm{Diag}(\hat{t}^{\,l}_{ik},\ k=1...K_j), \tag{9.9}$$

$$\hat{\mathbf{U}}^{m}_{i} = \mathrm{Diag}(\hat{\mathbf{U}}_{ij}^{\,m}), \qquad \hat{\mathbf{U}}_{ij}^{\,m} = \hat{u}_{ij}^{\,m} \times \mathbf{I}_{K_j}, \tag{9.10}$$

and $\mathbf{I}_{K_j}$ is a $(K_j \times K_j)$ identity matrix.[1)]

The fuzzy objective function Eq. 9.4 is similar to the objective functions used in Chapter 8, Eq. 8.3 and those proposed by Dunn (1974), Bezdek et al. (1981a), and others. The problem of minimizing fuzzy objective functions under restrictions has been well studied. Bezdek et al. (1981a) have proven the general theorem that estimators of memberships of the form (9.6) and (9.7) are necessary and sufficient for a strict local minimum of a fuzzy objective function of the form (9.4), in the nonsingular case. (The conditions (9.5), (9.6) and (9.7) are in general not sufficient for global optimality.) Eq. 9.5 shows that $\hat{\mathbf{b}}_i$ is well defined if rank$(\mathbf{X}'\hat{\mathbf{T}}^{l}_{i}\hat{\mathbf{U}}^{m}_{i}\mathbf{X})$ $\geqslant$ P. Singularity occurs only if any $\hat{G}_{ik} = 0$ or $\hat{D}_{ij} = 0$ (Dunn 1974), and necessitates tie-breaking rules which conform to the constraints (9.2) or (9.3).[2)]

Eq. 9.4 can be trivially minimized by degenerate solutions with for $i=i'$: $u_{i'j} = 1$, $t_{i'k} = 0$ ($j=1...n$, $k=1...K$) or vice versa, for which the constraint $\Sigma_k t_{ik} > 0$, or $\Sigma_j u_{ij} > 0$, respectively is not satisfied. Therefore, an appropriate normalization factor is used which produces weighted mean distance measures:[3)]

$$\hat{G}_{ik} = \Sigma_j \hat{u}_{ij}^{\,m}\, \hat{e}^{2}_{ijk} / \Sigma_j \hat{u}_{ij}^{\,m}, \qquad \hat{D}_{ij} = \Sigma_k \hat{t}^{\,l}_{ik}\, \hat{e}^{2}_{ijk} / \Sigma_k \hat{t}^{\,l}_{ik}, \tag{9.8a}$$

It can easily be observed that because of the normalization, for example $\hat{G}_{ik}$ can not become zero (resulting in $\hat{t}_{ik} = 1$) because of $\hat{u}_{ij}$ approaching zero (DeSarbo et al. 1984 used a similar normalization of a sum-of-squares measure to avoid degeneracies.)[4)]

GFCR solves the limitations of the traditional methods for benefit segmentation and market structuring discussed earlier. The lack of degrees of freedom at the individual level is solved by the simultaneous segmentation and estimation procedure. The predictive fit of the multiattribute models is maximized by the minimization of the weighted sum of squared residuals criterion J_{Rml}. Fuzzy segmentation and fuzzy market structuring are allowed for, while the degree of fuzziness can be influenced by the fuzzy weight parameters m and l. GFCR reveals the structure of the market in relation to benefit segments.

9.2.2 *The algorithm*

The iterative algorithm proposed entails the following steps (the validity of the algorithm using the modification given in Eq. 9.8a is demonstrated in section 9.3):

1. At the first step of the iterative process ($z=0$), initialize by fixing c, l and m. The starting matrices $\hat{\mathbf{U}}^{(0)}$ and $\hat{\mathbf{T}}^{(0)}$ are generated, e.g. from a uniform distribution, and normalized to satisfy the sum constraints (9.2) and (9.3).
2. Compute $\hat{\mathbf{b}}_i^{(z+1)}$, for $i=1...c$, according to Eq. 9.5
3. Calculate the new $\hat{\mathbf{U}}^{(z+1)}$ and $\hat{\mathbf{T}}^{(z+1)}$ from Eq. 9.6 and 9.7. $\hat{\mathbf{T}}^{(z)}$ is used in the expression for $\hat{\mathbf{U}}^{(z+1)}$, and $\hat{\mathbf{U}}^{(z)}$ in the expression for $\hat{\mathbf{T}}^{(z+1)}$.
4. Iterate between 2. and 3. until a prespecified change in J_{Rml} is met.

A tie-breaking rule similar to that given in Chapter 8 is applied in the case that for $j=j'$ at least one of the $\hat{D}_{ij'}=0$, or for $k=k'$ at least one of the $\hat{G}_{ik'}=0$. Let:

$$S = \{1,2...c\},$$

$$S_{j'} = \{i \in S \mid \hat{D}_{ij'}=0\},$$

$$S_{k'} = \{i \in S \mid \hat{G}_{ik'}=0\},$$

$n_{S_{j'}}$ and $n_{S_{k'}}$ denote the number of elements in $S_{j'}$ and $S_{k'}$ respectively.

Then the tie-breaking rules for u_{ij} or t_{ik} are:

$$\hat{u}_{ij'} = 0 \text{ for } i \in S\text{-}S_{j'}\text{; and } \hat{u}_{ij'} = 1/n_{S_{j'}} \text{ for } i \in S_{j'}, \tag{9.11}$$

$$\hat{t}_{ik'} = 0 \text{ for } i \in S\text{-}S_{k'}\text{; and } \hat{t}_{ik'} = 1/n_{S_{k'}} \text{ for } i \in S_{k'}. \tag{9.12}$$

A FORTRAN 77 program, GFCRCLUST, was developed that incorporates the above algorithm. The program operates idential to the FCRCLUST program described in 8.2.5, which it includes as a special case.

9.2.3 *A cross-validation procedure*

In this section, the cross-validation procedure proposed in 8.2.6 is extended to test both the clustering and predictive ability of GFCR (Of the available cross-validation procedures, sample-splitting, bootstrap, jackknife, simultaneous approaches, see Cooil et al. 1987 we use the sample-splitting method because it is easy to use and inexpensive.) In the proposed procedure a random sample, called the analysis sample, is selected to estimate memberships and preference weights in clusters. The results are used to classify the holdout (validation) sample and to predict the dependent variable in this sample. GFCR can both be cross-validated using a holdout sample of subjects that evaluated the same set of brands, or a hold out sample of products, evaluated by the same subjects. First, a procedure for validation with a holdout sample of subjects will be described that consists of six steps:

1. The sample of subjects is split randomly in two samples. One half is used as an analysis sample, the other half as a validation sample.
2. The analysis sample is submitted to GFCR (c, l and m are assumed to be known). This yields estimates of the coefficients, $\hat{\mathbf{b}}_i^{(a)}$, and memberships, $\hat{u}_{ij}^{(a)}$ and $\hat{t}_{ik}^{(a)}$, in the analysis sample.
3. The y variable in the validation set, $\mathbf{y}^{(v)}$, is predicted from the x variables, using the coefficients of each of the analysis sample segments in turn:

$$\hat{\mathbf{y}}_i^{(av)} = \mathbf{X}^{(v)} \hat{\mathbf{b}}_i^{(a)}, \tag{9.13}$$

where

$\hat{\mathbf{y}}_i^{(av)}$ = the vector of validation predictions in segment i,
$\mathbf{X}^{(v)}$ = the matrix of independent variables in the validation sample.

The cross-validatory percentage of variance explained, $\bar{R}_a^{2(av)}$ is calculated for each cluster.

4. The memberships $\hat{u}_{ij}^{(av)}$, that assign subjects in the validation sample to the analysis sample's segments are calculated according to Eq. 9.7, with:

$$\hat{D}_{ij}^{(av)} = \Sigma_k \hat{t}_{ik}^{(a)l}(y_{jk}^{(v)} - \hat{y}_{ijk}^{(av)})^2/\Sigma_k \hat{t}_{ik}^{(a)l} \tag{9.14}$$

5. The $\hat{u}_{ij}^{(av)}$ are used as a starting partition of a GFCR analysis of the validation sample, in which the brand memberships are fixed: $t_{ik} = \hat{t}_{ik}^{(a)}$, as memberships and preferences of a sample of new subjects, with respect to the same brands, are predicted. This yields estimates of coefficients, $\hat{\mathbf{b}}_i^{(v)}$, memberships, $\hat{u}_{ij}^{(v)}$, and predicted preferences $\hat{y}^{(v)}$, for the validation sample, as well as the percentage of variance explained by these estimates, $R_a^{2(v)}$.
6. Two validation statistics are calculated, assessing both the predictive and cluster validity of GFCR:
   $S_1 = R_a^{2(av)}/R_a^{2(v)}$, indicating the relative cross-validatory predictive accuracy, calculated for each cluster.
   $S_2 = r(\hat{u}_{ij}^{(v)}, \hat{u}_{ij}^{(av)})$, the cross-validity of the cluster solution, calculated for each cluster (r denotes the Pearson correlation coefficient).

The above procedure can similarly be applied to test the cross-validity of GFCR with respect to a hold out sample of brands. In step 1, the sample is partitioned into an analysis sample and a validation sample of brands. In step 2 the analysis sample coefficients and memberships are estimated, and in step 3 the preferences for the hold out set of brands are predicted from the analysis sample coefficients. In step 4 the cross-validatory brand memberships, $\hat{t}_{ik}^{(av)}$, are calculated from Eq. 9.6, with:

$$\hat{G}_{ik}^{(av)} = \Sigma_j \hat{u}_{ij}^{(a)m}(y_{jk}^{(v)} - \hat{y}_{ijk}^{(av)})^2/\Sigma_j \hat{u}_{ij}^{(a)m} \tag{9.15}$$

In many instances step 5, in which the holdout sample of brands is to be analyzed with GFCR, is not feasible, as only a small number of holdout brands will be available. Therefore, in step 6, the cross-validatory statistics will be expressed relative to the values of the sample of hold out brands obtained from the analysis of the total sample (analysis and validation sample combined):

$$S_1 = R_a^{2(av)}/R_a^2,$$

$$S_2 = r(\hat{t}_{ik}, \hat{t}_{ik}^{(av)}),$$

The results of a cross-validation study on a holdout sample of brands should be interpreted with caution, however, because of the small number of brands usually included, and the sensitivity of the results to the brands selected for the validation sample.

9.2.4 *Limitations of GFCR*

Convergence to local optima

As other partitioning clustering methods, GFCR may converge to local optima, depending upon the starting partition selected. A behavior of this sort is exacerbated by the absence of compact well separated clusters in the data (Dunn 1974). Whether or not local optima provide sufficient approximations is an empirical question. The problem can be overcome by having the algorithm started with different (random) initial partitions. An alternative solution is to have the algorithm started from a larger number of clusters and to work down to the desired number of clusters (Banfield and Bassil 1977). Alternatively, a more efficient initial partition may be used based, for instance, on a hierarchical clustering of preferences (Punj and Stewart 1983).

Selection of the number of clusters

The selection of the number of clusters (c) in partitioning clustering methods is currently a topic raised in the literature (e.g. Milligan and Cooper 1985). For GFCR we propose two criteria to aid in the selection of the number of clusters. The number of clusters can be determined from plots of the value of the criterion minimized, J_{Rml}, and the (adjusted) percentage of variance unexplained averaged across clusters, $1-\bar{R}_a^2$, against the number of clusters. The number of clusters selected is that number where the plots show an elbow, or level off.

Selection of the fuzzy weight parameters m and l

Just as in FCR, in GFCR the values of the fuzzy weight parameters m and l have to be chosen judgmentally. This problem is not specific to FCR or GFCR, but common to the class of fuzzy clustering algorithms to which it is related (Bezdek et al. 1981a, Hruschka 1986), and resembles the problem of selecting the value of r of the Minkowski r-metrics in nonmetric multidimensional scaling (Kruskal 1964). Values of m and l too close to 1 will result in nonoverlapping clusters, too large values will result in excessive overlap. Both types of solution are undesirable (Arabie et al. 1981) and have near-zero variance of those u_{ij} and t_{ik} that indicate substantial membership (i.e. u_{ij} and $t_{ik} > 1/c$). At intermediate values of m and l this variance of memberships is positive and may have an optimum. This suggests that plots of the standard error of $\hat{u}_{ij} > 1/c$ (SE_u) and $\hat{t}_{ik} > 1/c$ (SE_t) against a range of values of m and l may assist in the selection (see also section 8.2.3). The effects of m and l on the performance of GFCR will be investigated in section 9.3.

Significance testing

As was discussed in Chapters 7 and 8, the statistical tests commonly used in regression analysis (F-tests, t-tests) can not be used in clusterwise regression, as the distribution of the residual mean square within clusters is unknown, and the asymptotic properties do not hold (Wedel and Steenkamp 1989). The significance of the regressions within clusters can be examined with simplified Monte Carlo significance tests (Hope 1968, Wedel and Kistemaker 1989). In these procedures, the null hypothesis is to be rejected if the test criterion for the observed data (e.g. a t-value) exceeds the $M(\alpha/2)$ or the $M-M(\alpha/2)$ percentile of the test criterion calculated from analyses of a reference set, consisting of M-1 data sets generated from the observed data by random permutation of the dependent variable (α is the level of significance of the two-sided test and M is an integer). The power of the test increases with M (Hope 1968). (Alternatively, nonparametric resampling methods such as jackknife or bootstrap might be used to find the empirical distribution of the estimated coefficients (Cooil et al., 1987). However both of these methods are computationally more expensive than the Monte Carlo procedure.) It should be noted that in GFCR no distributional assumptions with respect to the dependent variable are necessary.

9.2.5 *Related procedures*

After its initial development by Späth (1979, 1981, 1982), clusterwise regression procedures for market segmentation have been proposed by Wedel and Kistemaker (1989, see Chapter 7), DeSarbo, Oliver and Rangaswamy (1989), DeSarbo and Cron (1988), Wedel and Steenkamp (1989, see Chapter 8), and Kamakura and Russell (1989). All of these methods yield a partition of one mode of a data set. The method described in Chapter 7 yields a hard partition of the consumer mode of three-way (consumers, products, variables) preference data. The method proposed by DeSarbo, Oliver and Rangaswamy (1989) yields an overlapping partition of the consumer mode of three-way data, a simulated annealing algorithm being used to maximize the variance accounted for. The method of DeSarbo and Cron (1988) yields a fuzzy partition of one mode of two-way (subjects, variables) data, the clusters being assumed to arise from a mixture of conditional normal distributions. An E-M algorithm is used to obtain maximum likelihood estimates of coefficients within clusters, and Bayes' rule is used to estimate posterior memberships of subjects in clusters. FCR (Chapter 8) yields a fuzzy partition of the consumer mode of three-way preference data, using the minimization of a distance criterion. It operates on a different principle than the mixture approach to clustering, as partial memberships are estimated from the data. Kamakura and Russell (1989) use the mixture approach to analyze three-way data (subjects, time, variables) of consumer choice. A mixture of multinomial distributions is assumed, and its parameters (memberships and price coefficients) are estimated using an E-M algorithm. Although the method deals with both segmentation and market structuring, as the other above methods, it yields a partition of the consumer mode of the data only.

Two-mode clustering procedures have been proposed by Hartigan (1975), DeSarbo (1982) and De Soete et al. (1984); the latter also review the literature on two-mode clustering. The methods are tailored to the clustering of row and column objects from two-mode rectangular proximities data, and do not allow for the simultaneous estimation of associations within clusters. Whereas the methods of Hartigan (1975) and DeSarbo (1982) allow for overlapping clusters, DeSoete et al. (1984) developed hierarchical procedures for estimating nonoverlapping clusters.

GFCR combines the estimation of regression models within clusters with the estimation of a two-mode (subjects and products) fuzzy partition of three-way data.

9.3 Monte Carlo analysis of performance

In order to assess the performance of the GFCR algorithm, a Monte Carlo simulation study was performed. Synthetic data sets were generated, with 2 or 4 clusters, 8 or 16 of products (K), 50 or 100 subjects (n), the level of error drawn randomly from a normal distribution with zero mean and a standard deviation (SD) of 0.05 or 0.10, and 3 or 5 x variables (P). These five factors were varied according to a fractional factorial 2^5 design in 8 trials (Cochran and Cox 1957). The responses, y^i_{jk}, of consumer j to brand k in cluster i were generated using the importance weights shown in Figure 9.1 (the x variables were drawn from a uniform distribution).

Figure 9.1 shows that, for the two-cluster data, subjects 1 to n/2 have two different preference functions: one for brands 1 to K/2 (cluster 1), the other for brands K/2 + 1 to K (cluster 2). Consequently, these subjects have a membership of 0.5 in each cluster. Subjects n/2 + 1 to n have one preference function for all brands, and have a membership of 1 in cluster 1. As a result, brands 1 to K/2 have a membership of 1 in cluster 1, and brands K/2 + 1 to K have a membership of 0.5 in each cluster.

For the four-cluster data, subjects 1 to n/2 have different preference functions for brands 1 to K/2 (cluster 1) and for brands K/2 + 1 to K (cluster 2), and so do subjects n/2 + 1 to n (clusters 3 and 4 respectively). This results in subjects 1 to n/2 having a membership of 0.5 in clusters 1 and 2, and subjects n/2 + 1 to n having a membership of 0.5 in clusters 3 and 4.

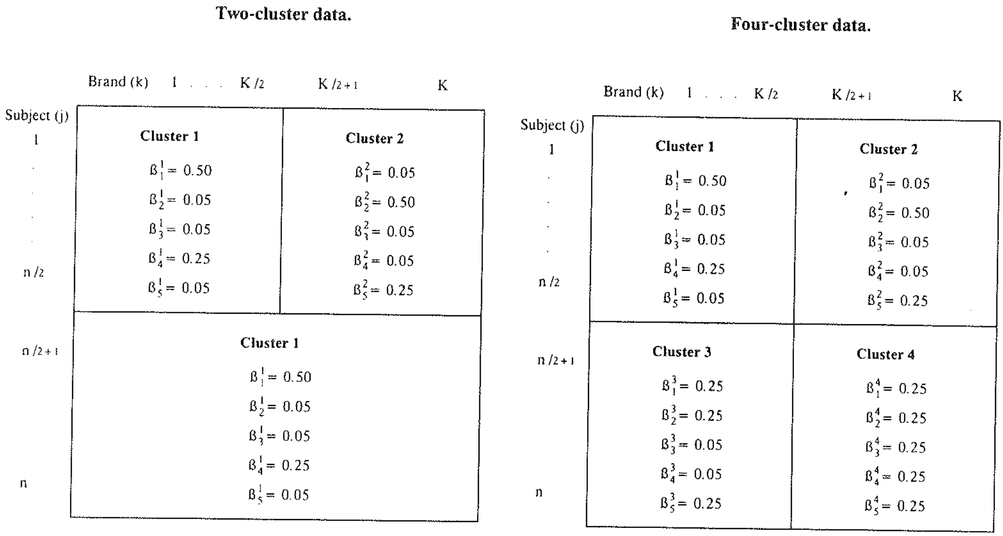

Figure 9.1 Importance weights within the two- and four-cluster data sets

Brands 1 to K/2 have a membership of 0.5 in clusters 1 and 3, and brands K/2+1 to K a membership of 0.5 in clusters 2 and 4. For $P=3$, the first three importance weights within each cluster were used in generating the data.

The eight data sets were analyzed with different values of the fuzzy weight parameters m and l, to investigate their effects on the performance of the algorithm. The fuzzy weights were varied according to a central composite design (Cochran and Cox 1957), based on a 2×2 factorial with levels 1.5 and 2.1 for both m and l, while the center point (1.8,1.8) and the star points (1.8,1.2), (1.8,2.4), (1.2,1.8), and (2.4,1.8) were added. This resulted in 9 replicate analyses of each data set. All 9 replicate analyses of each data set were started from the same random starting partition, which was chosen from a few trials to avoid local optima.

The GFCR solutions were evaluated according to a number of dependent measures:

1. $\bar{R}_a^2$, the average variance accounted for by GFCR across segments;

2. RMSE(b), the root mean squared error between the actual and estimated coefficients;
3. RMSE(u), the root mean squared error between the actual and estimated subject memberships;
4. RMSE(t), the root mean squared error between the actual and estimated product memberships;
5. $F(t) = \Sigma_i \Sigma_k t_{ik} \ln(t_{ik})$, the fuzziness of the solution with respect to the partition of products, normed by dividing by the value of the expression for $u_{ij} = 1/c$, so that it varies from 0 to 1, indicating increasing fuzziness;
6. $F(u) = \Sigma_i \Sigma_j u_{ij} \ln(u_{ij})$, the fuzziness of the solution with respect to the partition of subjects, normed by dividing by the value for $t_{ik} = 1/c$;
7. the number of iterations required (the upper limit was set to 100).

Table 9.1 shows the results of the GFCR analyses (m=l=2.1) of two of the eight synthetic data sets, one with two clusters and eight products, the other one with four clusters and sixteen products. Both data sets had an error level of 0.05 and included fifty subjects and three x variables. The recovery of memberships of both products and subjects (only a sample is shown) and the preference weights within clusters were quite accurate, as evidenced by RMSE(t), RMSE(u), and RMSE(b) respectively. Data reproduction, indicated by $\bar{R}_a^2$, was quite good for both data sets. The percentage of variance explained by the actual parameters was 90.1% for data set 1, and 87.4% for data set 2. GFCR attained about 98% of these values.

The 72 values (nine analyses of each of eight data sets) of the seven dependent measures in the Monte Carlo study were analyzed by linear regression, in which the main effects for the five factors (dummy coding), and linear effects, quadratic effects, and the interaction between m and l (coded in units of 0.1), were included. The interactions of the effects of m and l with the five factors were tested with sequential F-tests, to investigate if the effects of m and l on algorithm performance depended upon the factors of the study.

Table 9.1
Parameters recovered by GFCR (m = 2.1, l = 2.1) for two synthetic data sets

Data set 1

k	$\hat{t}_{1k}$	$\hat{t}_{2k}$	j	$\hat{u}_{1j}$	$\hat{u}_{2j}$	p	$\hat{b}_{1p}$	$\hat{b}_{2p}$
1	0.78	0.22	1	0.33	0.67	1	0.475	0.051
2	0.85	0.15	2	0.42	0.58	2	0.052	0.485
3	0.90	0.10	3	0.28	0.72	3	0.081	0.065
4	0.94	0.06	4	0.23	0.77			
5	0.46	0.54	5	0.42	0.58			
6	0.31	0.67	…	…	…			
7	0.23	0.77	26	0.99	0.01			
8	0.47	0.53	27	0.86	0.14			
			28	0.87	0.13			
			29	0.95	0.06			
			30	0.83	0.17			
			…	…	…			

Performance measures
$\bar{R}_a^2 = 87.9$, RMSE(b) = 0.019, RMSE(u) = 0.267, RMSE(t) = 0.247, F(u) = 0.686, F(t) = 0.729.

Data set 2

k	$\hat{t}_{1k}$	$\hat{t}_{2k}$	$\hat{t}_{3k}$	$\hat{t}_{4k}$	j	$\hat{u}_{1j}$	$\hat{u}_{2j}$	$\hat{u}_{3j}$	$\hat{u}_{4j}$
1	0.32	0.03	0.59	0.06	1	0.37	0.43	0.08	0.13
2	0.33	0.05	0.53	0.09	2	0.26	0.58	0.12	0.05
3	0.39	0.04	0.49	0.08	3	0.54	0.28	0.15	0.03
4	0.38	0.03	0.55	0.05	4	0.23	0.61	0.10	0.06
…	…	…	…	…	5	0.48	0.31	0.13	0.08
9	0.05	0.59	0.05	0.31	…	…	…	…	…
10	0.02	0.44	0.06	0.48	26	0.11	0.06	0.25	0.59
11	0.04	0.46	0.05	0.46	27	0.07	0.04	0.42	0.46
12	0.07	0.35	0.08	0.50	28	0.18	0.05	0.60	0.17
…	…	…	…	…	29	0.10	0.06	0.26	0.59
					30	0.08	0.06	0.30	0.57
					…	…	…	…	…

p	$\hat{b}_{1p}$	$\hat{b}_{2p}$	$\hat{b}_{3p}$	$\hat{b}_{4p}$
1	0.466	0.069	0.250	0.259
2	0.071	0.486	0.232	0.234
3	0.067	0.041	0.057	0.251

Performance measures
$\bar{R}_a^2 = 86.0$, RMSE(b) = 0.016, RMSE(u) = 0.247, RMSE(t) = 0.288, F(u) = 0.795, F(t) = 0.743

Table 9.2 reports the factor level means of the dependent measures. The effect reported for each factor are adjusted for the effects of all other factors, as well as fo the effects of m and l. For m and l the regression coefficients of the linear and quad ratic effects and their interaction are reported.

Table 9.2
Results of the Monte Carlo study on GFCR performance

Factors/ Levels	$\bar{R}^2_a$	RMSE(b)	RMSE(u)	RMSE(t)	F(u)	F(t)	Iter.
Variables							
3	75.2[a]	0.059	0.290	0.295	0.646[a]	0.554[a]	33.5
5	83.5	0.047	0.312	0.278	0.510	0.459	43.5
Segments							
2	77.0	0.038	0.310	0.258[a]	0.607	0.544	49.5[a]
4	81.8	0.068	0.292	0.315	0.550	0.468	27.5
Products							
8	85.4[a]	0.063	0.313	0.314[a]	0.496[a]	0.402[a]	40.5
16	73.4	0.042	0.289	0.259	0.661	0.610	36.6
Subjects							
50	82.8[a]	0.053	0.307	0.295	0.556	0.487	31.3[a]
100	76.0	0.053	0.295	0.278	0.601	0.525	45.8
Error							
0.05	92.4[a]	0.034[a]	0.268[a]	0.284	0.485[a]	0.413[a]	27.9[a]
0.10	66.4	0.072	0.334	0.288	0.672	0.600	49.2
m							
linear	1.1	-0.013	-0.098[b]	-0.010	0.188[b]	-0.034	2.2
quadratic	0.4	0.001	0.019[b]	-0.002	-0.044[b]	0.003	1.3
l							
linear	4.0	-0.013	-0.049	-0.087[b]	-0.005	0.144[b]	-3.0
quadratic	-0.4	0.001	0.004	0.019[b]	-0.001	-0.031[b]	-2.4
ml	1.8	0.005	0.016	0.008	0.013	0.021	4.4
F-fit[c]	33.3	5.1	6.8	5.9	43.9	59.2	5.3
F-lof[d]	0.8	0.1	2.1	0.1	0.3	1.0	0.0

[a] Significant difference between factor level means ($p<0.05$).
[b] Regression coefficient significantly different from zero ($p<0.05$).
[c] F-test for significance of the regression, df=10,61.
[d] F-test for lack-of-fit of the response surface of m and l, df=3,58.

From the sequential F-tests, none of the interactions of m and l with the design factors were significant. For none of the dependent measures the interaction between m and l was significant. All regression equations were strongly significant, and none of the analyses indicated a significant lack of fit of the quadratic response function of m and l. In 4 of the 72 analyses the algorithm did not converge within 100 iterations.

The percentage of variance accounted for by GFCR decreased with increasing numbers of products and subjects, increased with the number of x variables, and decreased with increasing error. $\bar{R}_a^2$ did not appear to be influenced by m and l across the range of values included in the study.

RMSE(b) increased significantly with increasing error level. The recovery of the preference weights within clusters was not affected significantly by the values of m and l chosen in the study. The average value of 0.053 indicated that the preference parameters were recovered quite accurately. (Note that neither the number of products nor the number of subjects affected the accuracy with which the parameters were estimated, contrary to the results found for FCR (Chapter 8); the range of these variables in the present study could have been too small to demonstrate the effects.)

RMSE(u) increased with the amount of error added to the y-variables in the synthetic data. RMSE(u) was significantly affected by m, both the linear and quadratic effects being significant. The significant quadratic effect indicates that there is an optimal value of m with respect to the recovery of the subject memberships. The average value of RMSE(u) was 0.301.

RMSE(t) was significantly affected by l, both the linear and the quadratic effects being significant, indicating that there is an optimal value of l with respect to recovery of the product memberships. Further, recovery of product memberships decreased for a larger number of segments, and improved for a larger number of products. The average value of RMSE(t) was 0.286.

The fuzziness of the partition of subjects and products, F(u) and F(t), increased significantly with increasing error, and with increasing numbers of products, and decreased with increasing numbers of x variables. F(u) is significantly affected by m, and F(t) by l, the significant quadratic coefficients indicating a curvature of the response surfaces.

The number of iterations increased with increasing number of subjects, and increasing error, while it decreased with increasing number of clusters. Neither m nor l affected the number of iterations.

In conclusion, the Monte Carlo analysis revealed several interesting findings. The parameters m and l affected parameter recovery of subject and product memberships, and the fuzziness of the partitions. With respect to the recovery of the memberships, the significance of the quadratic coefficients indicated that optimal values exist. Computational performance, data reproduction and the recovery of preference weights within clusters seem to be rather insensitive to m and l, at least within the range of values of m and l chosen in the study.

As the amount of error in the data increased, parameter recovery and data reproduction decreased, while computational requirements increased. Increasing the size of the sample (subjects and products) decreased data reproduction, and increased parameter recovery. (With respect to parameter recovery the majority of the effects, although consistent, were not significant, perhaps due to the range of values chosen in the study.) These findings agree with the findings for FCR (Chapter 8) and with traditional statistical estimation theory. Both data reproduction and parameter recovery of GFCR were shown to be quite satisfactory.

Some limitations to the Monte Carlo study should be noted. Emphasis was on the effects of m and l and their interactions with other factors on the performance of GFCR. Interaction effects between the other five factors could not be analyzed, and only two levels of each were specified.

9.4 Comparison with clusterwise regression and fuzzy clusterwise regression

To establish further the practical value of GFCR, it was compared to the clusterwise regression (CR) procedure of Wedel and Kistemaker (1989, see Chapter 7), which provides a nonoverlapping partition of subjects, and with the fuzzy clusterwise regression (FCR) procedure of Wedel and Steenkamp (1989, FCR, see Chapter 8), which provides a fuzzy partition of subjects. CR wa empirically compared to the two-stage procedure (section 7.3.3), FCR was empirically compared to the overlapping clustering method of DeSarbo et al. (1989) and to Hagerty's

(1985) optimal weighting (section 8.4). The three methods, CR, FCR, and GFCR, were applied to two of the synthetic data sets that were used in the Monte Carlo study in the previous section. Table 9.1 shows the results of the GFCR analyses of these data sets.

The number of clusters was varied from 2 to 5, for all three clusterwise regression procedures, to determine the number of clusters revealed by the respective methods. FCR was applied with m = 1.5 for data set 1 and m = 1.1 for data set 2 (these values were selected by the procedure outlined in section 8.2.3). Table 9.3 shows the percentage of variance not accounted for by the analyses.

Table 9.3
$1-\bar{R}_a^2$ against the number of clusters of the CR, FCR and GFCR analyses of two synthetic data sets

Number of clusters	Data set 1 CR	FCR	GFCR	Data set 2 CR	FCRG	FCR
1	34.1	34.1	34.1	43.7	43.7	43.7
2	25.0	25.4	12.1	41.7	41.9	35.7
3	22.6	25.1	6.8	41.3	41.6	21.2
4	20.9	25.6	3.1	40.4	40.1	14.0
5	20.3	25.5	0.8	39.7	40.1	16.6

As judged by the elbows in the plots of the percentage of variance not accounted for against the number of clusters, for data set 1 all three methods indicated a two-cluster solution, which is the actual number of clusters present in the data (for GFCR the three-cluster solution could also have been appropriate, but one cluster was recovered twice in this solution). For data set 2, however, both CR and FCR indicated a two-cluster solution, while GFCR indicates the correct four-cluster solution to be optimal. As both CR and FCR entail a grouping of consumers only, this result is not surprising as data set 2 in fact contains two clusters along the consumer mode. Tables 9.4 and 9.5 show the parameter estimates and the performance measures of CR, and FCR respectively. The RMSE(u) and

RMSE(b) for these analyses were calculated with respect to the actual partition of the consumer mode only.

For data set 1, the performance of CR and FCR was much less than that of GFCR, as evidenced by $\bar{R}_a^2$, RMSE(u), and RMSE(b). Not only estimated GFCR the importance weights within clusters more accurately, it also estimated the memberships of subjects within segments better. FCR recovered both memberships and importance weights better than CR. It is clear that although both CR and FCR indicated the two-cluster solution to be optimal, they only estimated the partition of the consumer mode. Both methods recovered the average values of the importance weights in consumer segments, across the product clusters within these segments. The average values of the importance weights in consumer segments 1 and 2, averaged across product clusters, were: $\mathbf{b}'_1 = (0.5, 0.05, 0.05)$, $\mathbf{b}'_2 = (0.275, 0.275, 0.05)$.

For data set 2, GFCR also clearly outperformed both FCR and CR, as evidenced by $\bar{R}_a^2$, RMSE(b) and RMSE(u). FCR and CR were unable to recover the product clusters, nor could they estimate the subject memberships as accurately as GFCR. FCR estimated the importance weights of the two segments (which are the averages across the two brand clusters within the two segments) as accurately as GFCR estimated the importance weights of the two-way cluster structure. The average values of the importance weights within consumer segments 1 and 2 were: $\mathbf{b}'_1 = (0.25, 0.25, 0.05)$, $\mathbf{b}'_2 = (0.25, 0.25, 0.15)$.

Even if for the CR and FCR analyses of data set 2 the four-cluster solution was chosen (the number of clusters actually present in the data) GFCR outperformed both methods. The performance of CR on data set 2 (c = 4) was:

$\bar{R}_a^2 = 59.6$, RMSE(b) = 0.101, RMSE(u) = 0.442, F(u) = 0.

The performance of FCR on data set 2 (c = 4) was:

$\bar{R}_a^2 = 59.4$, RMSE(b) = 0.077, RMSE(u) = 0.388, F(u) = 0.433.

Table 9.4

Parameters recovered by CR for two synthetic data sets

Data set 1

j	$\hat{u}_j$	$\hat{u}_{2j}$	p	$\hat{b}_p$	$\hat{b}_{2p}$
1	1	0	1	0.183	0.477
2	0	1	2	0.355	0.076
3	1	0	3	0.060	0.060
4	1	0			
5	1	0			
-	-	-			
26	0	1			
27	0	1			
28	0	1			
29	0	1			
30	0	1			
-	-	-			

Performance measures

$\bar{R}_a^2 = 75.0$, RMSE(b) = 0.082, RMSE(u) = 0.354, F(u) = 0

Data set 2

j	$\hat{u}_{1j}$	$\hat{u}_{2j}$	p	$\hat{b}_{1p}$	$\hat{b}_{2p}$
1	1	0	1	0.299	0.251
2	1	0	2	0.241	0.242
3	0	1	3	0.039	0.152
4	1	0			
5	0	1			
-	-	-			
26	0	1			
27	0	1			
28	0	1			
29	1	0			
30	0	1			
-	-	-			

Performance measures

$\bar{R}_a^2 = 58.3$, RMSE(b) = 0.045, RMSE(u) = 0.447, F(u) = 0

Table 9.5

Parameters recovered by FCR ($m = 1.5$) for two synthetic data sets

Data set 1

j	u_{1j}	$\hat{u}_{2j}$	p	$\hat{b}_{1p}$	b_{2p}
1	0.67	0.33	1	0.173	0.467
2	0.20	0.80	2	0.366	0.083
3	0.93	0.07	3	0.064	0.063
4	0.92	0.08			
5	0.92	0.08			
~	~	~			
26	0.00	1.00			
27	0.02	0.98			
28	0.04	0.96			
29	0.00	1.00			
30	0.05	0.95			
~	~	~			

Performance measures

$\bar{R}_a^2 = 74.6$, RMSE(b) = 0.060, RMSE(u) = 0.291, F(u) = 0.404

Data set 2

j	$\hat{u}_{1j}$	$\hat{u}_{2j}$	p	$\hat{b}_{1p}$	$\hat{b}_{2p}$
1	0.55	0.27	1	0.290	0.260
2	0.96	0.24	2	0.247	0.236
3	0.67	0.25	3	0.040	0.146
4	0.96	0.23			
5	0.35	0.25			
~	~	~			
26	0.07	0.93			
27	0.00	1.00			
28	0.04	0.96			
29	0.52	0.48			
30	0.03	0.97			
~	~	~			

Performance measures

$\bar{R}_a^2 = 58.1$, RMSE(b) = 0.015, RMSE(u) = 0.383, F(u) = 0.443

(The performance measures were calculated with respect to the actual four clusters present in data set 2.) Parameter recovery for the four-cluster solution of both methods is worse than that of the corresponding two-cluster solution. FCR also shows a better parameter recovery than CR for the four-cluster solution.

Summarizing, both CR and FCR are unable to recover the two-way cluster structure that was present in the two synthetic data sets. GFCR outperforms FCR and CR both in data reproduction and parameter recovery, while FCR performed somewhat better than CR in recovering the partition of the subject mode.

9.5 Application to data on preferences for butter and margarine brands

9.5.1 *Data*

GFCR was used to reanalyze data on consumer preferences for butter and 12 margarine brands in the Netherlands (Steenkamp and Meulenberg 1986). In the market definition both margarine brands and butter are included, because they are substitutes for use on bread and for frying or baking. In a nationwide sample of 535 subjects, all of whom were the main purchasers of food in the household, data were collected by interviews at home. Subjects were asked to classify brands (including butter) into five or less groups of similar brands. The similarity data were aggregated across respondents, and MDSCAL was applied to the aggregate data. The stress of the four-dimensional solution was 0.012, which compares favorably to the results reported by Klahr (1969). The scores of the brands on the MDS dimensions are shown in Table 9.6. The first three dimensions had a clear interpretation: exclusiveness, vegetable component, and fitness for multiple purposes. The fourth dimension was more difficult to interpret, but was associated with the type of packaging (stick versus tub). The four dimensions were to be related to stated preferences.

Preference data were obtained by asking the subjects to rank the brands in ascending order of preference.

Table 9.6

Scores of margarine brands and butter on the four MDS dimensions exclusiveness (EX), vegetable component (VE), fitness for multiple purposes (FI), and packaging (PA)

Brand		EX	VE	FI	PA
1 Brio	(br)	-0.231	0.833	-0.432	0.222
2 Becel	(be)	0.389	1.078	-0.096	-0.345
3 Bona	(bo)	0.247	0.329	0.286	-0.755
4 Morgen	(mo)	0.569	-0.556	-0.652	0.083
5 Gouda's glorie	(gg)	-0.240	-0.244	0.555	-0.251
6 Leeuwezegel	(le)	-0.316	-0.296	0.551	0.273
7 Zeeuws meisje	(zm)	-0.304	-0.447	0.562	0.207
8 Remia	(re)	-0.621	-0.181	-0.367	0.036
9 Butter	(bu)	1.835	-0.179	-0.228	0.430
10 Blue band	(bb)	-0.024	-0.348	0.677	0.284
11 Wajang	(wa)	-0.564	0.301	-0.556	0.342
12 AH margarine	(ah)	-0.400	-0.489	0.157	0.161
13 Sun	(su)	-0.325	0.216	-0.471	-0.691

Additional information was obtained on urbanization, annual household income, age, socioeconomic status and the psychological variable 'locus of control', measured on a Dutch version of Rotter's (1966) scale.

9.5.2 *Results and implications*

For the analyses a random sample of 267 subjects was drawn, the other subjects were to serve as a holdout sample for cross- validation. For the purpose of illustration, an extensive procedure for selecting m and l will be employed. GFCR was performed for a number of values of m and l, varied according to the central composite design with the center at (1.8,1.8) the same that was used in section 9.3, and according to a 3×3 factorial design with levels (1.1,1.2,1.3) for both parameters. For each of the resulting solutions the standard errors of product (SE_t) and subject (SE_u) memberships larger than 0.5 were calculated, and analyzed by linear regression with linear and quadratic effects of m and l as independent variables. Linear and quadratic effects of m on SE_u, and of l on SE_t were significant. In

Figure 9.2 the average values of SE_u are plotted against m, the average values of SE_t against l. From the plot $m = l = 1.5$ were chosen for further analyses. (The analyses suggest that in applications m may be set equal to l, which facilitates the search for optimal values.)

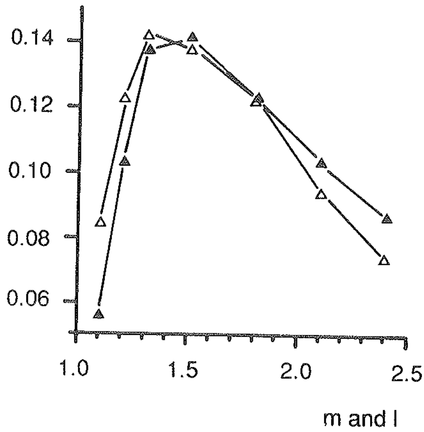

Figure 9.2 Plot of average values of SE_u (△) and SE_t (▲) against m and l for the butter and margarine data

From the plot $1-\bar{R}_a^2$ against the number of clusters, the three-cluster solution appeared the most appropriate, while the plot of J_{Rml} also be justified the two-cluster solution (Figure 9.3).

The three-cluster solution was inspected, but did not provide insights additional to the two-cluster solution. GFCR was run five more times for $c = 2$, with a random initial partition. The same solution was recovered four times. This solution was the final solution decided upon.

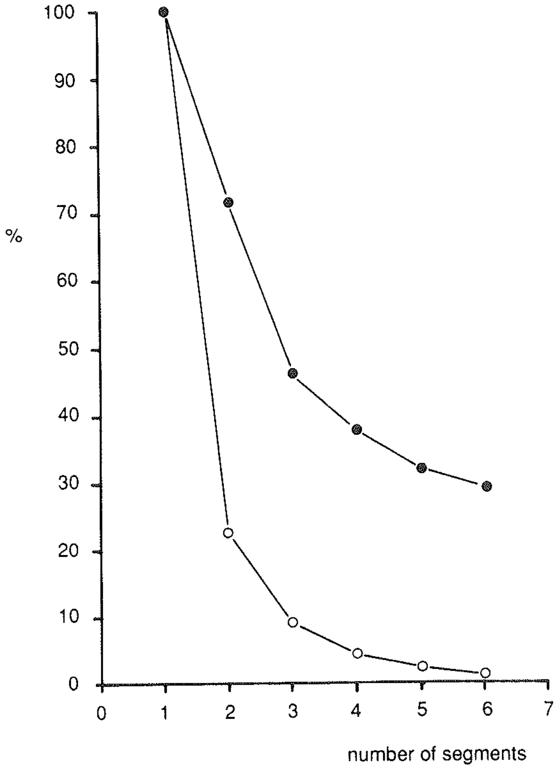

Figure 9.3 Plot of J_{Rml} (o) and $1\text{-}\bar{R}_a^2$ (●) against the number of clusters for the butter and margarine data

The results of the two-cluster solution of GFCR and the total sample regression analysis are shown in Table 9.7. The two criteria proposed by Hauser and Urban (1977) support the internal validity of the solution: the preference models within clusters fit substantially better than the total sample regression, and the preference weights show considerable differences between clusters.

Table 9.7

Results of the two-cluster solution of the GFCR analysis (m = 1.5, l = 1.5) of data on preferences for butter and margarine brands

Attribute	Cluster 1	Cluster 2	Total sample
Exclusiveness	0.496	2.992	2.136
(-2.37 - 2.66)[a]	(3.50)	(46.95)	(23.35)
Vegetable component	1.960	2.470	2.191
(-2.12 - 1.82)	(13.87)	(21.42)	(16.66)
Fitness multiple purposes	4.633	1.830	3.126
(-2.43 - 2.84)	(47.43)	(11.37)	(23.22)
Packaging	0.617	1.500	1.275
(-1.94 - 2.70)	(4.23)	(9.02)	(7.63)
Average memberships			
Products	0.510	0.490	1.000
Subjects	0.467	0.533	1.000
R_a^2	0.423	0.509	0.251

[a] 2.5 and 97.5 percentiles of the distribution of t in the reference set.
[b] t-values are given in parenthesis.

In both clusters all coefficients are significant by the Monte Carlo test procedure. Table 9.8 contains the brand memberships and the predicted preferences within the clusters.

It is important to note that a high membership for a brand does not imply that the brand competes in the cluster in question. The degree of membership of a brand in a cluster reflects the degree to which preferences for the brand, which may be either high or low, are based on the benefits that are considered important by consumers in the segment, as is indicated by the segments' preference function (Table 9.7). Depending on the consumers' perception of the benefits (Table 9.6), brands with a high membership may rate high or low on preference in the segment in question. Therefore, marketing implications of the solution should be based on an evaluation of brand memberships, preferences (Table 9.8) and importance weights (Table 9.7), in relation to the perceptual structure (Table 9.6).

Table 9.8
Brand memberships ($\hat{t}_{ik}$) and predicted preferences ($\hat{y}_{ik}$) of the two-cluster GFCR solution for the butter and margarine data

Brand k	Cluster 1 $\hat{t}_{1k}$	Cluster 1 $\hat{y}_{1k}$	Cluster 2 $\hat{t}_{2k}$	Cluster 2 $\hat{y}_{2k}$	Total $\hat{y}_k$
1 Brio	0.43	6.8	0.57	8.0	7.5
2 Becel	0.15	8.8	0.85	10.3	9.6
3 Bona	0.59	8.8	0.41	8.1	8.4
4 Morgen	0.84	3.4	0.16	6.4	5.3
5 Gouda's glorie	0.54	9.0	0.46	6.5	7.6
6 Leeuwezegel	0.42	9.2	0.58	6.9	7.9
7 Zeeuws meisje	0.48	8.9	0.52	6.5	7.6
8 Remia	0.74	4.9	0.26	4.2	4.4
9 Butter	0.01	7.0	0.99	12.4	10.6
10 Blue band	0.77	9.8	0.23	7.9	8.9
11 Wajang	0.73	5.1	0.27	5.7	5.3
12 AH margarine	0.31	6.9	0.69	5.3	6.0
13 Sun	0.64	4.8	0.36	4.8	4.6

Brands are assumed to compete within a cluster when the same preference function underlies high consumer preference: high membership of a brand and high preference for a brand in a cluster indicates a strong position. Low membership and high preference indicates a relatively strong position that can be reinforced by a change in competitive structure towards the brands in the cluster. High membership and low preference indicates a weak position, and low membership and low preference indicates a very weak position. Consequently, the implications of the analyses for marketing strategy are differentiated, even for different brands within the same cluster.

In Figure 9.4 the 13 brands are plotted on the two dimensions that most strongly differentiate the clusters: exclusiveness and fitness for multiple purposes. The cluster preference functions are inserted into the plot.

In cluster 1, fitness for multiple purposes is the most important benefit sought. The more common brands, which are perceived to be fit for multiple purposes, have substantial memberships and high preferences in cluster 1: Bona, Gouda's glorie, Leeuwezegel, Zeeuws meisje and Blue band. These brands compete in segment 1. The brands Morgen, Remia, Wajang and Sun have high memberships but low preferences, caused by low perceived fitness for multiple purposes. These brands can increase their appeal in this segment by enhancing their perceived fitness for multiple purposes. Supplementary information revealed that especially Sun is perceived unfit for baking and frying. If this is perception is incorrect communication strategies can be employed, but otherwise possibilities for product modification should be assessed. The analyses revealed a very weak position for the brand Morgen.

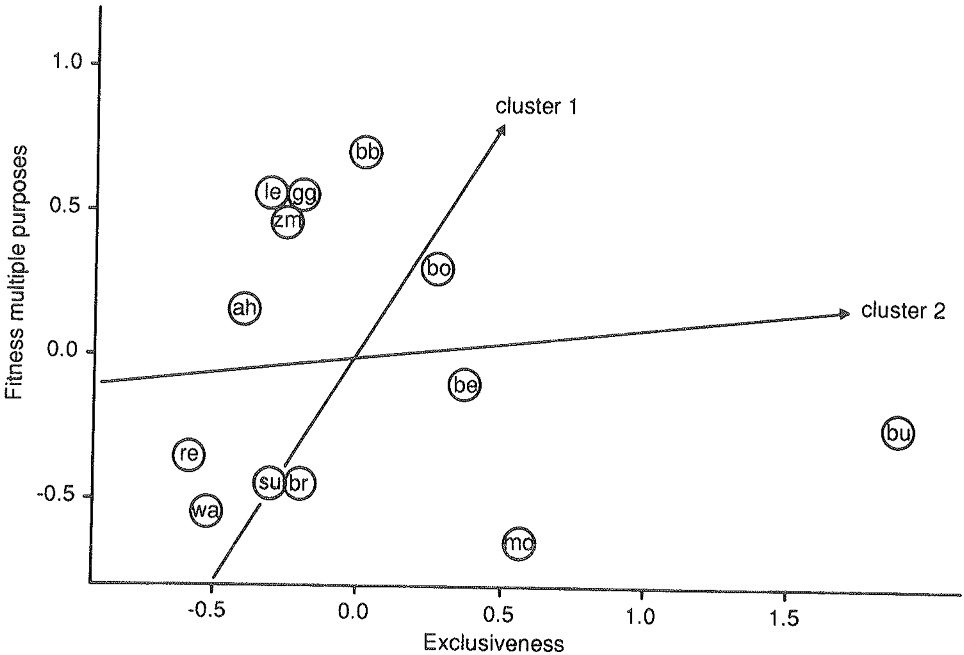

Figure 9.4 Plot of the positions of the margarine brands and butter and the cluster preference functions with respect to the MDS dimensions exclusiveness and fitness for multiple purposes (See Table 9.6 for explanation of the symbols)

In cluster 2, high preferences are found for the 'special' brands, which rate high on exclusiveness or vegetable component. Brio, Becel, Bona and butter are strong competitors. Gouda's glorie, Leeuwezegel and AH margarine have high memberships, but predicted preferences are low. If management wishes to focus upon segment 2, these brands can increase their appeal by enhancing consumer perceptions on the exclusiveness dimension, for example by advertising or packaging, whereas product modification to increase the proportion of vegetable component might be advantageous as well. AH margarine could focus on segment 1 by increasing its perceived fitness for multiple purposes.

Becel, Bona and Blue band have high preferences in both segments, relatively independent of their membership in the clusters, and thus cover the whole market and compete in both segments. Remia, Wajang and Sun have a weak position in both segments, which is evidenced by the low preferences. It might be advantageous for these brands to focus on segment 1, and improve perceived fitness for multiple purposes. On the other hand, modification of the brands to further increase the vegetable component (Wajang already scores reasonably high on that dimension) will increase consumer preferences in segment 2. Given the growing awareness of the relationship between food and health, the latter option might be the most attractive.

A number of limitations to the empirical application should be noted. First, the analyses concern data collected in 1983, and the butter and margarine market in the Netherlands has changed since. The translation of the results of the segmentation study does not only depend on the validity of GFCR (which will be investigated in section 9.5.3) but also on the validity of the MDS solution and its interpretation. The main distinction in benefits revealed was in exclusiveness and fitness for multiple purposes. Although they have a different profile of perceived attributes, butter and Becel were revealed to compete in the same segment, in which exclusiveness and vegetable component are traded off. The vegetable component dimension, on which a number of brands are positioned, did not discriminate between segments. This may be hypothesized to be attributable to the relatively uniform judgment of its importance by all consumers in different usage situations.

The extent to which the revealed segments can be effectively reached through communication and distribution channels depends on the extent to which consumers that have a high segment membership can be profiled with demographic and socioeconomic characteristics. Psychographic variables will provide further clues for the development of advertising messages or product modifications. The segments revealed by GFCR were characterized in terms of consumer descriptors in a second step of the analyses. A dummy variable regression was performed of the log of the ratio of subject memberships between clusters 1 and 2, on the variables sex, age, urbanization, socioeconomic status, income and locus of control ($R_a^2 = 6.3\%$). Women and subjects of a higher socioeconomic class had 2.0 and 2.1 times higher memberships respectively in cluster 2 than in cluster 1 ($p < 0.05$). Although the percentage of variance explained by the regression model is low, these relationships are consistent and support the validity of the segment solution. Only a limited number of consumer characteristics was assessed in the present study. Other lifestyle variables, more specifically related to the behavior towards buying butter and margarine might have revealed stronger relationships with segment membership. However, because of lack of data, these analyses could not be performed.

9.5.3 *Cross-validation results*

The analyses of the previous section were performed on a random sample of 267 subjects. The other subjects served as a holdout sample to assess the cross-validity of GFCR according to the procedure outlined in section 9.3.2. The values of $m = l = 1.5$ and $c = 2$ determined in the previous section were used for the analyses. Table 9.9 shows the cross-validation results.

The average validation statistic $\bar{S}_1$ was 0.939 which indicated that cross-predictive accuracy attained over 90% of the predictive fit of the GFCR analysis of the validation sample, which is quite satisfactory. (Note that the percentage of variance accounted for by cross-prediction, $R_a^{2(av)}$, is higher in each of the clusters than the predictive fit of a regression model fitted on the validation sample itself, for which $R_a^2 = 0.247$).

The statistic S_2, the correlation of cross-predicted and estimated subject memberships indicates a very close correspondence.

Table 9.9

Results of the GFCR cross-validation study of a holdout sample of subjects for the butter and margarine data

Cluster	$R_a^{2(av)}$	$R_a^{2(v)}$	S_1	S_2
1	0.324	0.324	0.999	0.988
2	0.475	0.540	0.879	0.988

In order to investigate the ability of GFCR to predict memberships and preferences for a new set of brands, a sample of 3 brands (Bona, Leeuwezegel and Wajang) was selected, and the analysis set, consisting of the preferences of 267 subjects for the 10 remaining brands, was analyzed with GFCR ($m=l=1.5$, $c=2$). Table 9.10 shows the cross-validation results.

Table 9.10

Results of the GFCR cross-validation study of a holdout sample of brands for the butter and margarine data

	Cluster 1		Cluster 2	
	$\hat{t}_{1k}^{(av)}$	$\hat{y}_{1k}^{(av)}$	$\hat{t}_{2k}^{(av)}$	$\hat{y}_{2k}^{(av)}$
Bona	0.55	8.5	0.45	7.5
Leeuwezegel	0.43	9.6	0.57	6.8
Wajang	0.61	5.3	0.39	6.0

Cluster	$R_a^{2(av)}$	R_a^2	S_1	S_2
1	0.305	0.379	0.806	0.991
2	0.153	0.154	0.993	0.996

The cross-predicted memberships and preferences (Table 9.10) for the hold-out brands are reasonably close to the results of the total sample analysis (Table 9.8), as is evidenced by the statistic S_2 (the average correlation of memberships is 0.994, but note that there is only 1 df).

The average validation statistic $\bar{S}_1$ was 0.900, which shows that the cross-predictive accuracy for the holdout sample of brands is quite satisfactory. On the whole, the results for the holdout brands are somewhat less convincing than the results for the holdout sample of subjects shown in Table 9.9. This might be the result of the particular set of brands selected for the holdout sample. The cross-validation of brands performed is illustrative, but generalization of the results should be viewed with caution. The analyses are based on small numbers of brands in the analysis and validation samples, and are sensitive to the specific brands included in the samples. A jackknife validation procedure might produce better results.

In spite of some of the drawbacks related to the data-splitting procedure used (underutilization of available information, larger prediction errors at the validation stage, and validation results that are dependent upon the particular split selected; see Cooil et al. 1987), it may be concluded that the cross-validity of GFCR, with respect to both prediction and classification of holdout samples of subjects and brands, was satisfactory. Cooil et al. (1987) argued that data-splitting is inefficient because it underutilizes available information. Consequently, the results obtained with this procedure will be conservative estimates of the validity of the method. The cross-validity of GFCR should be further investigated in empirical applications, where especially for brands the (time-consuming) jackknife or bootstrap procedure might be used.

9.6 Conclusions

This chapter has presented a method for simultaneous benefit segmentation and market structuring. The validity of the algorithm was supported by the analyses of synthetic data in a Monte Carlo study, and by comparisons with clusterwise regression and fuzzy clusterwise regression. In GFCR the competitive structure of the market is revealed in relation to benefit segments. Insight is given into the underlying causes of market structures and segments, on the basis of multi-attribute models of consumer behavior, as was illustrated in the empirical application to consumer preferences for margarine brands and butter. The validity of the resulting solution was supported by cross-validations on holdout samples of consumers as well as brands.

It was demonstrated that the segments revealed by GFCR can be made accessible by relating them to consumer descriptors. Although the results were consistent, the percentage of variance explained was rather small. Variables that are more specifically related to the buying behavior studied might have demonstrated stronger relationships. On the other hand, the methods could be extended to allow for a simultaneous description of segments with consumer characteristics.

There are a number of limitations to GFCR. When a larger number of clusters is imposed upon the data, the unexplained variance approaches zero, and the estimates of the importances become unstable due to near-zero residual variance. This problem becomes more salient for small values of m and l (especially in applications where the x matrix is the same for all subjects), because such values will result in observations receiving zero weight in the analyses, which enhances problems with multicollinearity.

Another problem related to GFCR is the choice of the fuzzy weight parameters. In the Monte Carlo analysis of the performance of the method it was shown, however, that parameter recovery and predictive fit were rather insensitive to m and l, within the (broad) range of values investigated. The fuzzy weight parameters influenced the memberships of brands and subjects significantly, and thereby the fuzziness of the resulting solutions. Solutions of GFCR obtained from the same data with different values of m and l are consistent. A heuristic procedure, proposed for determining m and l, guards against nonoverlapping and excessively overlapping solutions (both of which are undesirable). The analyses on empirical data suggest that in practice m may be set equal to l, which facilitates the search for optimal values. The proposed algorithm could be extended to incorporate an iterative search for optimal m and l. However such an extension is left for further study.

10 GENERAL DISCUSSION AND CONCLUSIONS

10.1 Bases and methods

In Part 1 of this work, the abundance of bases and methods that have been used to identify segments were reviewed. In Chapter 3, benefits and benefit importances were identified as one of the most effective bases for segmentation. Other available bases should not be discarded, however. In segmentation designs, each may serve according to its own merits in relation to the purpose of the study. Demographic and socioeconomic bases can be used to demonstrate the accessibility of segments, psychographics are especially appealing for developing advertising messages and new products, and preferences and intentions find their value in their ability to predict behavior.

Not only the use of different bases will lead to different segments being revealed, the same holds for the use of different methods. In Chapter 4, from a review of the literature, the post-hoc clustering methods were shown to deal more adequately with the complexity of markets than the a-priori methods. Fuzzy clustering methods are especially attractive because they allow for partial membership of consumers in segments, thus alleviating the less realistic assumption that consumers belong to only one segment. The potential of the predictive pattern techniques for segmentation research was demonstrated, as they combine the post-hoc identification of segments with prediction of a dependent measure of interest, be it preferences, intentions or behavior.

The clusterwise regression methods for segmentation of consumer markets presented in this work, CR, FCR and GFCR, fall into the class of predictive pattern techniques. In the approaches, benefits, their importances, and preferences or intentions are the key segmentation variables. The methods, being linked to the multiattribute models developed for describing consumers' attitude structure, explicitly maximize the accuracy with which preferences or intentions are predicted. In the application of CR, the analyses were restricted to one usage situation, while FCR and GFCR explicitly allow for consumers to have partial membership in more than one segment. Other segmentation variables were used in the approaches according to their major strengths: demographic, socioeconomic and psychographic variables were used to demonstrate accessibility, to enhance actionability, and to add to the understanding of why differences in behavior between segments occur.

10.2 Clusterwise regression methods: assumptions involved

The methods presented in Part 2 of this work find their applications within the behavioral school of segmentation research, due to their concern with multiattribute models, preference prediction, and benefit segmentation. In the microeconomic school, the methods proposed by Grover and Srinivasan (1987, 1989) and Kamakura and Russell (1989) deal with fuzzy consumer segmentation. Both are latent-class type of models. Whereas the methods of Grover and Srinivasan (1987, 1989) are predominantly descriptive, the Kamakura and Russell (1989) approach entails a demand response model.

The clusterwise regression methods of DeSarbo, Oliver and Rangaswamy (1989) and DeSarbo and Cron (1988) were developed in a broader context than for segmentation alone, but, due to the distributional assumptions (normality), they are applicable predominantly within the behavioral school. In the development of their clusterwise regression method, DeSarbo and Cron (1988) also followed a latent-class approach.

Very recently, a number of clusterwise regression methods have been added to this list. Kamakura and Agrawal (1990) proposed a mixture of multinomial logit models for simultaneously classifying a sample of consumers into benefit segments and estimating their random utility functions. The method was developed for the analysis of rank-order preferences collected in conjoint measurement. De Soete and DeSarbo (1990) developed a latent-class probit model for the analysis of binary pick-any-out of n data[1)].

The methods for fuzzy clusterwise regression presented in Chapters 8 and 9 fall within the same realm, although operating on a somewhat different principle. Whereas in the mixture approaches consumers are assigned to segments with a-posteriori calculated probabilities, the fuzzy methods presented in this work assume that consumers actually belong to more than one segment, and the partial segment memberships are estimated explicitly as parameters in the model. As compared to the mixture approaches the fuzzy methods have the disadvantages of having to select the values of the fuzzy weights, and of not being based on maximum likelihood properties. On the other hand in FCR and GFCR, the fuzzy weight parameters provide the users with flexibility with respect to the degree of partitioning to be obtained, and no distributional assumptions on the dependent variable have to be made. In the conditional mixture approaches the type of relationship between dependent and independent variables depends on the distribution that is assumed for the dependent variable, whereas in FCR and GFCR any appropriate transformation of the dependent variable may be employed to linearize the relationships.

In applications, selection of a method should be based upon a careful consideration of its representation of the structure of the data in relation to the underlying assumptions. Assumptions that are possibly involved are:

1. the number of segments is known
2. the segments are nonoverlapping
3. the relevant set of brands is identical across subjects
4. the coefficients for all subjects within a segment are identical
5. the distribution of the dependent variable is known
6. the preference functions are homogeneous across products

The hierarchical methods proposed by Kamakura (1988) and Ogawa (1987) require assumptions 2 to 6; the method presented in Chapter 7 (Wedel and Kistemaker 1989) requires assumptions 1, 2, and 4 to 6; DeSarbo, Oliver and Rangaswamy (1989) require 1, and 3 to 6; DeSarbo and Cron (1988), Kamakura and Russell (1989), Kamakura and Argawal (1990), and De Soete and DeSarbo (1990) require 1, 3, 5, and 6; FCR (Chapter 8, Wedel and Steenkamp 1989) requires 1 and 6; GFCR (Chapter 9) requires only assumption 1. Additionally, the latter two methods also require the determination of the fuzzy-weight parameters.

Another criterion that may be used in the choice of one of these methods are the computational requirements: little is known of the relative performances of the methods in this respect, although in FCR and GFCR there is an additional amount of computations involved in the selection of the fuzzy-weight parameters and the

Monte Carlo significance tests. Kamakura (1990) has presented empirical evidence that the E-M algorithm on which many of the mixture approaches are based is relatively inefficient, and proposed a more efficient algorithm.

The clusterwise regression approaches thus incorporate the major features of segmentation research that have been identified as potentially effective in determining the segment structure of markets. The methods could be extended further to provide a simultaneous description of segments in terms of consumer characteristics, or to deal explicitly with usage situations.

10.3 Substantive results

The major segmentation bases used in this work were benefit importances. Substantive results of the proposed methods were obtained in the empirical studies, all concerning foods. The five salient benefits for food products used in the applications comprised taste, wholesomeness, convenience, price and prestige. The benefits were operationalized somewhat differently in different studies, being derived by a variety of methods, such as direct rating, factor analysis, and multidimensional scaling. The clusterwise regression procedures were applied in conjunction with conjoint analysis as well. This 'standard set' of benefits for food products also plays a role in the formation of perceived quality. Perceived quality mediates the effect of product attributes on preference and is traded off against price in preference formation. It was shown that, for various groups of food products not only different consumer segments exist with respect to the tradeoff of the salient benefits in the formation of preferences, but that also different groups of consumers place different relative importances on perceived quality and price.

The major differences in the segment structure between different food markets that were revealed concerned predominantly the relative magnitudes of the importances and the number of segments that were identified.

With respect to store image, the standard set of five attributes that was used was: product quality, price, service quality, store atmosphere and assortment. Consequently, the segments that were identified in the application place different importances on these attributes.

In all of the applications, the relationships of consumer descriptors with segment membership were weak, irrespective of the statistical method used to assess these relationships (contingency tables, cross-tabulation, discriminant analysis, regression, and partial least squares). The salient general consumer descriptors (sex, age income, socioeconomic status) generally exhibited weaker relationships than product-specific variables (nutritional knowledge, health locus of control, involvement). These considerations suggest that, dependent on the purpose of the study, the above sets of attributes and product-specific variables should be considered for future investigations into the segmentation of food preferences and (food) store image.

The application of clusterwise regression procedures is not restricted to the perceptions and preferences data that were used in the empirical sections of this work. Other fields of marketing research where they could find their application are the analysis of scanner data from consumer or retail panels, direct marketing, industrial marketing, management research, and services marketing. Potential fields of application outside of marketing research are in political science, psychology, social science, and nutrition.

10.4 Implications for the development of marketing strategy

Segmentation imposes a partition of the market and identifies submarkets, the needs of which can be addressed more precisely. Segmentation research aims at identifying of a segment structure of markets, and thus plays an important role in the development of a marketing strategy.

The application of proper segmentation methods, with an adequate commitment to the underlying assumptions, is crucial with respect to the number and type of segments that are identified. Not only the use of different bases may lead to different segment structures being revealed, much the same holds for the application of different segmentation methods. Whereas initially segmentation research questions had to be forced into the available statistical methods, more and more statistical methods have been developed that are tailored to specific segmentation problems. The clusterwise regression methods developed in this work were specifically designed to solve the problems traditionally encountered in benefit segmentation. That is why, that in the application of these methods, there is an optimal match between the structure of the data and the method and its underlying assumptions.

In the applications of the clusterwise regression methods, it was exemplified how the results can indicate possible marketing strategies. In the cases where market segments can be distinguished, the methods provide insight into the structural attractiveness of segments, which has to be evaluated in relation to the firm's resources in the development of marketing strategy. The methods give a picture of the strength of competition between products/brands within segments, in terms of high consumer preferences. Insight is provided into the underlying causes of competitive structures, using the multiattribute model of attitude formation. The competitive advantages or disadvantages of products/brands are signaled, and the product-market fit is revealed in terms of the product attributes that are considered important in segments of the market. Opportunities for brands can be identified, and clues for product modification or communication strategies are provided.

The extent to which segments can be reached through distribution or communication channels depends on the extent to which these segments can be profiled with demographic, socioeconomic, and psychographic variables. Whereas communicating or distributing to nonoverlapping segments seems conceptually simpler than to fuzzy segments, this is not the case. When segments are fuzzy, consumer memberships in segments can be meaningfully associated with consumer descriptors, as was shown in various applications in this work. Consequently, the fuzzy segments are equally accessible as nonoverlapping segments. By designing a communication strategy that appeals to a fuzzy target segment, or by choosing distribution or communication channels that match the target segments profile, some consumers (with high memberships) are reached more effectively than others (with lower memberships). In fact, the development of a marketing strategy by a marketing manager on the basis of revealed fuzzy segments proceeds exactly as on the basis of nonoverlapping segments.

The recent research thrust devoted to the development and application of the fuzzy approaches to segmentation suggests the onset of a turbulent change in segmentation theory and practice. The availability of the new clusterwise regression methods alleviates the problems that have long been associated with the traditional procedures in both the behavioral and microeconomic school, related to the estimation of preference model parameters, or local measures of demand, at the individual level. The full potential of the clusterwise regression approach in a large number of areas of segmentation research is still to be exploited.

Wind (1978) called for new analytic methods that place fewer demands on the consumer, and for approaches that provide a new conceptualization of the segmentation problem. The clusterwise regression methods, among which the methods presented in this work, are believed to meet Wind's requirements, and to be a valuable adjunct to existing marketing segmentation approaches.

Chapter 3

1) Wilkie and Cohen (1977) provided another viewpoint. They developed a framework to place segmentation studies into a behavioral perspective. The system is based upon the expectancy value theory, and classifies the bases used for segmentation into: (1) general person descriptors, (2) psychographics, (3) desired values, (4) brand perceptions, (5) attitudes, preferences and intentions, and (6) purchase behavior. The system is reductive in that a segmentation base on a certain level is influenced by all lower-level bases. Dickson (1982) included segmentation according to the usage situations in Wilkie and Cohen's framework.

Chapter 4

1) Q-type factor analysis (FA), for example, is frequently mentioned as a method for segmentation, especially in psychographic research. The method is not specially suited for segmentation purposes, as it does not yield some grouping of the consumer sample. Q-type factors are not interpretable as (overlapping) clusters. The number of factors is not related to the number of clusters present, and the identification of homogeneous segments on the basis of FA is subjective and complex, especially when more than two factors are extracted (Stewart 1981). Furthermore, the number of factors that can be identified is necessarily less than the number of variables. A modification of FA, called Linear Typal Analysis (LTA), was suggested for market segmentation by Darden and Perreault (1977). They argued that, from a theoretical point of view, the Q-type factors can be seen as prototype consumers. Respondent profiles are linear combinations of these pure types. LTA specifically addresses the problem that whereas consumer profiles are characterized by their shape, elevation and scatter, Q-type FA only considers shape. In LTA a fixed number K is added to the matrix of cross products before factoring, the level of K being varied to determine the optimal value. FA and LTA operate from a different

concept than segmentation methods. Whereas they assume the existence of prototype consumers, true consumers being linear combinations of the prototypes, segmentation assumes the identifiability of homogeneous groups of consumers.

Other methods that have been suggested for segmentation research but will not be considered in this chapter are Multidimensional Scaling, Procrustes Analysis, Linear Structural Relations and Canonical Correlation (see e.g. Wind 1978).

2) Linear modeling techniques are specifically suited for predictive segmentation. Regression, log-linear models, multinomial logit and probit models are special cases of the class of generalized linear models (McCullagh and Nelder, 1989). These GLMs entail the specification of a linear model, of a distribution of the dependent variable (Normal, Poisson, Binomial, Multinomial, Gamma, Inverse Gaussian) and of a link function (identity, log, logit, probit, inverse), to relate the linear model to the expectation of the dependent variable. An iterative reweighted least-squares procedure is applied to obtain maximum likelihood estimates of the parameters of the linear model, of which (asymptotic) standard errors are provided. Models with only a specification of the mean and variance instead of the full distribution of the dependent variable - so called quasi likelihood models - can also be handled with this procedure.

3) It should be noted that the optimal variable weighting procedures (e.g. De Soete et al. 1985, DeSarbo et al. 1984) were not considered in this comparison, and that these procedures are less sensitive to the presence of variables on which the clusters are not distinguished.

Chapter 5

1) It can be argued that some of these approaches presented in the sequel also belong to the behavioral school of segmentation research. We will classify these approaches into the microeconomic stream on the basis of their concern with the prediction of demand (or purchase behavior).

Chapter 6

1) Among the noncompensatory models used in marketing literature are the conjunctive, disjunctive and lexicographic models. In the conjunctive model a consumer considers a brand only if it meets certain minimum standards on key attributes, in the disjunctive model only brands are considered that exceed acceptability levels on one or a few attributes, in the lexicographic model brands are compared on the most important attribute (Kotler 1988). However, even if the decision process is more complex, the compensatory models produce good predictions (Green and Srinivasan 1978, Wilkie and Pessemier 1972).

2) The different terms value, importance, salience and evaluation that are used in different models of attitude structure are not synonymous; see for discussions on this topic, for example, Cohen, Fishbein and Ahtola 1972, Sheth 1972, Talarzyk 1972, Holbrook and Hubert 1975, and Curry and Menasco 1983).

Chapter 7

1) The assumption of independently Normal-distributed error terms is not tenable when rank order preferences are used as a dependent variable. However, violations of this assumption are likely to be of less influence for larger numbers of products, as preference models are fitted across subjects and products. Moreover, the asymptotic properties of the ML estimates do not apply, irrespective of the distribution of the error terms, and Monte Carlo test procedures should be used for significance testing (section 7.2.2).

2) If full-rank preferences are assessed and the relevant sets of alternatives differ across consumers, the assumption of a common vector of preference parameters can not hold. Consequently, the method can only be applied to the analysis of rank order data if the relevant set is identical for all individuals.

3) This frequently used two-stage procedure, related to principal components regression, has the disadvantage that the perceptual dimensions and their number are determined on the basis of the representation of attribute ratings only. Consequently, these dimensions need not be optimally related to preferences, as variation in product attributes not accounted for by the perceptual dimensions may be a good predictor. The method of partial least squares overcomes this problem (Martens and Martens 1986), but the size of the present data set exceeded the capacity of the available program.

4) As the clusterwise regression model is additive between two overlapping segments, actually a third segment is identified. Subjects with membership to two segments form another segment with a regression model that differs from subjects belonging to only one of the segments.

Chapter 8

1) Hereafter we will use the term preference weights, and not refer to part-worths separately.

2) The assumption of independent error terms in the preference models is not met when full-rank preferences are assessed. However, violations of this assumption are likely to be of less influence for larger numbers of products, as preference models are fitted across subjects and products. The asymptotic properties of ML estimates do not apply, irrespective of the dependence of the error terms, and Monte Carlo test procedures should be used (see section 2.3).

3) Alternatively, or in addition, R_a^2, the average percentage of variance accounted for across segments, might be used in the selection of m.

4) We do not mean to imply that the models of DeSarbo et al. (1989) and Wedel and Kistemaker (1989) are special cases or constrained forms of our model, but rather that FCR can be applied to a wider variety of segmentation problems.

5) A computer program, PRDICT, additional to the fuzzy-c varieties clustering program FCVPC, has been developed for this purpose (personal communication, R.W. Gunderson, Department of Mathematics, Utah State University, Logan, Utah).

6) Traditional benefit segmentation using inferred weights is not even possible as only one observation is available for each subject.

Chapter 9

1) The augmented Lagrangian only constrains the sum of memberships to 1, and there could thus conceivably be negative numbers. However, it may be observed from Eq. 9.8a that $\hat{u}_{ij}$ is positive if $\hat{t}_{ik}$ is positive and visa versa. Consequently, when the iterative procedure proposed in section 9.2.2 is started with positive values, the memberships are enforced being between 0 and 1.

2) In this case a hard partition of subjects or brands or both is obtained, which results, in observations receiving zero weight in Eq. 9.5, which may lead to rank $(\mathbf{X}\hat{\mathbf{T}}_i^l \hat{\mathbf{U}}_i^m \mathbf{X}) < P$. This problem is more pregnant in applications where the x matrix is identical for all subjects such as in conjoint analysis, and if l is close to 1.

3) Alternatively, a different specification of J_{Rml} can be chosen, using an additive instead of multiplicative weight function of the memberships: $t_{ik}^l + u_{ij}^m$. It follows that Eq. 9.5, 9.6 and 9.7 apply with: $\hat{D}_{ik} = \Sigma_j \hat{e}^2_{ijk}$, and $\hat{D}_{ij} = \Sigma_k \hat{e}^2_{ijk}$. Due to the additivity of the weight function, solutions with in cluster i all $u_{ij} = 1$ and $t_{ik} = 0$ or visa versa do not result in $J_{Rml} = 0$. This alternative was not investigated exhaustively.

4) When the predictor variables have the same values for all subjects (for example when aggregate MDS or factor analysis dimensions are obtained, or in conjoint analysis), it can be shown that that the $\mathbf{b}_i$ obtained from Eq. 9.5 by regression

of $\mathbf{y}$ on $\mathbf{X}$, weighted with $\mathbf{T}_i^l\ \mathbf{U}_i^m$, are identical to the coefficients of the regression of $\mathbf{y}_i^*$ on $\mathbf{X}^*$, weighted with $\mathbf{T}_i^{*l}\ \mathbf{U}_i^{*m}$. The (K×1) vector $\mathbf{y}_i^*$ has elements: $y^*_{ik} = \Sigma_j u_{ij}^m\ y_{ijk}/\Sigma_j u_{ij}^m$,
$\mathbf{X}^*$ denoting the aggregate (K×P) x-matrix, $\mathbf{T}_i^{*l}$ the (K×K) diagonal matrix Diag(t_{ik}^l; k = 1...K), and $\mathbf{U}_i^{*m}$ the (K×K) diagonal matrix with elements $u_{ik}^{m} = \Sigma_j u_{ij}^m$. $\mathrm{Var}(\hat{\mathbf{b}}_i) = \mathrm{var}(\hat{\mathbf{b}}_i^*) s^2/s^{*2}$, where s^2 and s^{*2} are the respective residual variances of the regressions of $\mathbf{y}$ on $\mathbf{X}$, and of $\mathbf{y}^*$ on $\mathbf{X}^*$ respectively. Consequently, in this situation GFCR can be applied to data vectors of length K instead of length N, which reduces the computational time required.

Chapter 10

1) These data arise e.g. from a situation where a number of subjects are presented a set of n products/brands, and each subject picks any number of these objects which he intends to buy within some designated time period.

12.1 Appendix A: Derivation of the estimators of FCR

The objective is to find the estimates of $\mathbf{b}_i$, and $\mathbf{U}_i$, that minimize J_{Rm}:

$$J_{Rm} = \Sigma_i \ \Sigma_j \ \Sigma_k \, u_{ij}^m \, e_{ijk}^2 , \tag{A1}$$

under the constraint: $\Sigma_i u_{ij} = 1$. The Lagrangian is:

$$L = \Sigma_i \ \Sigma_j \ \Sigma_k \, u_{ij}^m \, e_{ijk}^2 - \mu(\Sigma_i \ u_{ij} - 1), \tag{A2}$$

with μ constant. Setting the derivatives with respect to the parameters equal to zero yields:

$$\delta L/\delta u_{ij} = m u_{ij}^{m-1} \ \Sigma_k \, e_{ijk}^2 - \mu = 0, \tag{A3}$$

$$\delta L/\delta \mathbf{b}_i = -2X' \, \mathbf{U}_i^m \, \mathbf{b}_i + X' \, \mathbf{U}_i^m \, \mathbf{y} = 0, \tag{A4}$$

$$\delta L/\delta \mu = \Sigma_i \, u_{ij} - 1 = 0. \tag{A5}$$

Solving A3 for u_{ij} yields:

$$\hat{u}_{ij} = (\mu/\Sigma_k m \ \hat{e}_{ijk}^2)^{1/(m-1)} \tag{A6}$$

By summing over i in A6, and using the constraints A5:

$$\mu = \{ \Sigma_i 1/(m \, \hat{e}_{ijk}^2)^{1/(m-1)} \}^{1-m} \tag{A7}$$

Eq. 8.5 follows from substitution of A7 in A6. From A4 it can be seen that the estimates of $\mathbf{b}_i$ are the ordinary weighted-least-squares estimates with weight matrix $\hat{\mathbf{U}}_i^m$. Solving A4 for $\hat{\mathbf{b}}_i$ yields Eq. 8.4.

Appendix B: Derivation of the estimators of GFCR

The objective is to find the estimates of $\mathbf{b}_i$, $\mathbf{U}_i$, and $\mathbf{T}_i$, that minimize J_{Rml}:

$$J_{Rml} = \Sigma_i \Sigma_j \Sigma_k u_{ij}^m t_{ik}^l e_{ijk}^2, \tag{B1}$$

under the constraints: $\Sigma_i u_{ij} = 1$ and $\Sigma_i t_{ik} = 1$. The Lagrangian is:

$$L = \Sigma_i \Sigma_j \Sigma_k u_{ij}^m t_{ik}^l e_{ijk}^2 - \mu(\Sigma_i u_{ij} - 1) - \lambda(\Sigma_i t_{ik} - 1) \tag{B2}$$

with μ and λ constant. Setting the derivatives with respect to the parameters equal to zero yields:

$$\delta L/\delta u_{ij} = m u_{ij}^{m-1} \Sigma_k t_{ik}^l e_{ijk}^2 - \mu = 0, \tag{B3}$$

$$\delta L/\delta t_{ik} = l t_{ik}^{l-1} \Sigma_j u_{ij}^m e_{ijk}^2 - \lambda = 0, \tag{B4}$$

$$\delta L/\delta \mathbf{b}_i = -2\mathbf{X}'\mathbf{T}_i^l \mathbf{U}_i^m \mathbf{b}_i + \mathbf{X}'\mathbf{T}_i^l \mathbf{U}_i^m \mathbf{y} = 0, \tag{B5}$$

$$\delta L/\delta \mu = \Sigma_i u_{ij} - 1 = 0. \tag{B6}$$

$$\delta L/\delta \lambda = \Sigma_i t_{ik} - 1 = 0. \tag{B7}$$

Solving B3 and B4 for u_{ij} and t_{ik} respectively yields:

$$\hat{u}_{ij} = (\mu/\Sigma_k m\, \hat{t}_{ik}\, \hat{e}_{ijk}^2)^{1/(m-1)} \tag{B8}$$

$$\hat{t}_{ik} = (\lambda/\Sigma_j m\, \hat{u}_{ij}\, \hat{e}_{ijk}^2)^{1/(l-1)} \tag{B9}$$

By summing over i in B8 and B9, and using the constraints B6 and B7:

$$\mu = \{ \Sigma_i 1/(m\, t_{ik}^l\, \hat{e}_{ijk}^2)^{1/(m-1)} \}^{1-m} \tag{B10}$$

$$\lambda = \{ \Sigma_i 1/(m\, u_{ij}^m\, \hat{e}_{ijk}^2)^{1/(l-1)} \}^{1-l} \tag{B11}$$

Eq. 9.7 and 9.6 follow from substitution of B10 in B8, and B11 in B9 respectively. From B5 it can be seen that the estimates of $\mathbf{b}_i$ are the ordinary weighted-least-squares estimates with weight matrix $\hat{\mathbf{T}}_i^{l}\ \hat{\mathbf{U}}_i^{m}$. Solving B5 for $\hat{\mathbf{b}}_i$ yields Eq. 9.5.

REFERENCES

Addelman, S. (1962), "Orthogonal main effect plans for asymmetrical factorial experiments", *Technometrics* 4, 21-46.

Adler, I. and D. Kafry (1980), "Capturing and clustering judges' policies", *Organizational Behavior and Human Performance* 25, 384-394.

Alvey, et al. (1977), *GENSTAT a general statistical program*, Rothamsted experimental station: Harpenden, England.

Andriessen, J.H.T.M. (1972), "Interne of externe beheersing", *Nederlands Tijdschrift voor Psychologie* 27, 173-198.

Arabie, P., (1977), "Clustering representations of group overlap", *Journal of Mathematical Sociology* 5, 113-128.

Arabie, P., and J.D. Carrol (1980), "MAPCLUS: "A mathemical programming approach to fitting the ADCLUS model", *Psychometrika* 45, 211-235.

Arabie, P., J.D. Carroll, W.S. DeSarbo and J. Wind (1981), "Overlapping clustering: a new method for product positioning", *Journal of Marketing Research* 18 (August), 310-317.

Assael, H. (1970), "Segmenting markets by group purchasing behavior: an application of the AID technique", *Journal of Marketing Research* 7 (May), 153-158.

Assael, H. and A.M. Roscoe Jr. (1976), "Approaches to market segmentation analysis", *Journal of Marketing* 40, 67-76.

Baker, M.J. (1988), *Marketing Strategy and management*, MacMillan Education, New York.

Banfield, C.F. and L.C. Bassil (1977), "A transfer algorithm for non-hierarchical classification", (Algorithm AS113), *Applied Statistics* 26, 206-210.

Basford, K.E. and G.J. McLachlan (1985), "The mixture method of clustering applied to three-way data", *Journal of Classification* 2, 109-125.

Bass, F.M. (1974), "The theory of stochastic preference and brand switching", *Journal of Marketing Research* 11 (February), 1-20.

Bass, F.M. and W.W. Talarzyk (1972), "An attitude model for the study of brand preference", *Journal of Marketing Research* 9 (February), 93-96.

Bass, F.M., D.J. Tigert, and R.T. Lonsdale (1968), "Market segmentation: group versus individual behavior", *Journal of Marketing Research* 5 (August), 264-270.

Bass, F.M., and W.L. Wilkie (1973), "A comparative analysis of attitudinal predictions of brand preference", *Journal of Marketing Research* 10 (August), 262-269.

Bearden, W.O. (1977), "Determinant attributes of store patronage: downtown versus outlying shopping centers", *Journal of Retailing* 53 (2), 15-22.

Beckwith, N.E. and M.W. Sasieni (1976), "Criteria for market segmentation studies", *Management Sciences* 22 (8), 892-908.

Beldo, L.A. (1966), "Market segmentation and food consumption", *Harvard Business Review* 184-195.

Belk, R.W. (1975), "Situational variables and consumer behavior", *Journal of Consumer Research* 2 (December), 157-164.

Ben-Akiva, M. and S.R. Lerman (1985), *Discrete choice analysis theory an application to travel demand*, The MIT Press, Cambridge, Massachusetts.

Berk, K.N. (1984), "Validating regression procedures with new data", *Technometrics* 26, 331-339.

Bezdek, J.C. (1974), "Numerical taxonomy with fuzzy sets", *Journal of Mathematical Biology* 1, 57-71.

Bezdek, J.C., C. Coray, R. Gunderson and J. Watson (1981a), "Detection and characterization of cluster substructure. I.Linear structure: Fuzzy c-lines", *SIAM Journal on Applied Mathematics* 40, 339-357.

Bezdek, J.C., C. Coray, R. Gunderson and J. Watson (1981b), "Detection and characterization of cluste substructure. II.Fuzzy c-varieties and convex combinations thereof", *SIAM Journal on Applied Mathematics* 40, 358-372.

Blashfield, R.K. and M.R. Aldenerfer (1978), "The literature on cluster analysis", *Multivariate Behavioral Research* 13 (July), 271-295.

Blattberg, R.C., T. Buesing, P. Peacock and S.K. Sen (1978), "Identifying the deal prone segment" *Journal of Marketing Research* 15 (August), 369-377.

Blattberg, R.C., T. Buesing and S.K. Sen (1980), "Segmentation strategies for new national brands" *Journal of Marketing* 44, 59-67.

Blattberg, R.C. and S.K. Sen (1974), "Market segmentation using models of multidimensional purchasin behavior", *Journal of Marketing* 38 (October), 17-28.

Blattberg, R.C. and S.K. Sen (1976), "Market segments and stochastic brand choice models", *Journal o Marketing Research* 13 (February), 34-45.

Blozan, W. and P. Prabhaker (1984), "Notes on aggregation criteria in market segmentation" *Journal o Marketing Research* 21, 332-335.

Bottenberg, R.A. and R.E. Christal (1978), "Grouping criteria - a method which retains maximun predictive efficiency", *The Journal of Experimental Education* 36 (4), 28-34.

Boyd, H.W. and W.F. Massy (1972), *Marketing Management*, Harcourt, Brace, Jovanovich, New York.

Breiman, L., J.H. Friedman, R.A. Olshen and C.J. Stone (1984), *Classification and regression trees*, Wadsworth International Group, Belmont, CA.

Brookhouse, K.J., R.M. Guion and M.E. Doherty (1986), "Social desirability response bias as one sourc of the discrepancy between subjective weights and regression weights", *Organizational Behavior and Human Decision Processes* 37, 316-328.

Calantone, R.J. and A.G. Sawyer (1978), "The stability of benefit segments", *Journal of Marketing Research* 15 (August), 395-404.

Carmone, F.J. and P.E. Green (1981), "Model misspecification in multiattribute parameter estimation" *Journal of Marketing Research* 18 (February), 87-93.

Carroll, J.D. (1972), "Individual differences and multidimensional scaling", In: R.N. Shepard, A.R Romney, S.B. Nerlove (eds.), *Multidimensional scaling*, Seminar Press: New York, 104-154.

Carroll, J. and P. Arabie (1983), "INDCLUS: an individual differences generalization of the ADCLU model and the MAPCLUS algorithm", *The Psychometric Society* 48 (2), 157-69.

Carroll, D.J. and J. Chang (1970), "Analysis of individual differences in multidimensional scaling via an n way generalization of Eckart-Young decomposition", *Psychometrika* 35, (September), 61-68.

Christal R.E. (1968), "JAN: a technique for analyzing group judgment", *The Journal of Experimental Education* 36 (4), 24-27.

Clarke, M.R.B. (1982), "The Gauss-Jordan sweep operator with detection of collinearity", (Algorith AS178), *Applied Statistics* 31, 166-168.

Claycamp, H.J. and W.F. Massy (1968), "A theory of market segmentation", *Journal of Marketing Research* 5 (November), 388-394.

Cochran, W.G. and G.M. Cox (1957), *Experimental designs*, John Wiley and sons, New York.

Cohen, J.B., M. Fishbein and O.T. Ahtola (1972), "The nature and uses of expectancy-value models in consumer attitude research", *Journal of Marketing Research* 9 (November), 456-460.

Cooil, B., R.S. Winer and D.L. Rados (1987), "Cross-Validation for prediction", *Journal of Marketing Research* 24 (August), 271-279.

Cormack, R.M. (1971), "A review of classification", *Journal of the Royal Statistical Society* 134, 321-353.

Cox, D.R. and D.V. Hinkley (1974), *Theoretical Statistics*. Chapman Hall, London.

Cronbach, L.J. (1971), "Test validation", In: R.L. Thorndike (ed), *Educational measurement*, American council on education, Washington D.C.

Currim, I.S. (1981), "Using segmentation approaches for better prediction and understanding from consumer mode choice models", *Journal of Marketing Research* 18 (August), 301-309.

Curry, D.J. and M.B. Menasco (1983), "On the separability of weights and brand values: issues and empirical results", *Journal of Consumer Research* 10 (1), 83-85.

Daganzo, C. (1979), *Multinomial probit, the theory and its application to demand forecasting*, Academic Press, London.

Darden, W.R. and W.D. Perreault (1977), "Classification for market segmentation: an improved linear model for solving problems of arbitrary origin", *Management Science* 24 (3), 259-271.

Day, G.S., A.D. Shocker and R.K. Srivastava (1979), "Customer-oriented approaches to identifying product-markets", *Journal of Marketing* 43, 8-19.

Day, N.E. (1969), "Estimating the components of a mixture of normal distributions", *Biometrika* 56 (3), 463.

DeSarbo, W.S. (1982), "GENNCLUS: new models for general nonhierarchical clustering analysis", *Psychometrika* 47 (4), 449-476.

DeSarbo, W.S., J.D. Carroll and L.A. Clarck (1984), "Synthesized clustering: A method for amalgamating alternative clustering bases with differential weighting of variables", *Psychometrika* 49, 57-78.

DeSarbo, W.S. and W.L. Cron (1988), "A maximum likelihood methodology for clusterwise linear regression", *Journal of Classification* 5, 249-282.

DeSarbo, W.S. and V. Mahajan (1984), "Constrained classification: the use of a priori information in cluster analysis", *Psychometrika* 49, 57-78.

DeSarbo, W.S., R.L. Oliver and A. Rangaswamy (1989), "A simulated annealing methodology for clusterwise linear regression", *Psychometrika* 54, 707-736.

De Soete, G., and W.S. DeSarbo, (1990)," A latent class probit model for analyzing pick any/N data", *Psychometrika, in press*.

De Soete, G., W.S. DeSarbo and J.D. Carroll (1985), "Optimal variable weighting for hierarchical clustering: an alternating least-squares algorithm", *Journal of Classification* 2, 173-192.

De Soete, G., W.S DeSarbo, G.W. Furnas and J.D. Carroll (1984), "The estimation of ultrametric and path length trees from rectangular proximity data", *Psychometrika* 49 (3), 289-310.

Dhalla, N.K. and W.H. Mahatoo (1976), "Expanding the scope of segmentation research", *Journal of Marketing* 40, 34-41.

Dickson, P.R. (1982), "Person-situation: segmentation's missing link", *Journal of Marketing* 46, 56-64.

Dickson, P.R. and J.L. Ginter (1987), "Market segmentation, product differentiation, and marketin strategy", *Journal of Marketing* 51 (April), 1-11.

Dillon, W.R. and M. Goldstein (1984), *Multivariate analysis, methods and applications*, Wiley, New York.

Donnelly, J.H. (1970), "Social character and acceptance of new products", *Journal of Marketing Research* 7 (February), 111-113.

Dorfman, R. and P.O. Steiner (1954), "Optimal advertising and optimal quality", *American Economi Review* 44, 826-836.

Doyle, P. and I. Fenwick (1974), "How store image affects shopping habits in grocery chains", *Journal o Retailing* 50 (4), 39-52.

Doyle, P. and I. Fenwick (1975), "The pitfalls of AID analysis", *Journal of Marketing Research* 12 (November), 408-413.

Doyle, P. and P. Hutchinson (1976), "The identification of target markets", *Decision Sciences* 7, 152 161.

Draper, N.R. and H. Smith (1976), *Applied regression analysis*, John Wiley and Sons, New York.

Dunn, J.C. (1974), "A fuzzy relative of the ISODATA process and its use in detecting compact wel separated clusters", *Journal of Cybernetics* 3, (3), 32-57.

Elrod, T. and R.S. Winer (1982), "An empirical evaluation of aggregation approaches for developin market segments", *Journal of Marketing* 46, 65-74.

Engel, J.F., H.F. Fiorillo and M.A. Cayley, (eds.) (1972), *Market Segmentation: Concepts and Applications*, Holt, Rinehart and Winston, New York.

Evans, F.B. (1959), "Psychological and objective factors in prediction of brand choice", *Journal o Business* 17, 340-369.

Fishbein, M. (1967), *Attitude and the prediction of behavior: a behavior theory approach to the relations between beliefs about an object and the attitude toward the object*, Readings in attitude theory and measurement, Wiley, New York.

Fishbein, M. and I. Ajzen (1975), *Belief, attitude, intention, and behavior: an introduction to theory and research*, Addison-Wesley, New York.

Frank, R.E. (1967), "Is brand loyalty a useful basis for market segmentation, *Journal of Advertising Research* 7, 27-33.

Frank R.E. (1972), 'Predicting new product segments", *Journal of Advertising Research* 12 (3), 9-13.

Frank, R.E. and P.E. Green (1968), "Numerical taxonomy in marketing analysis: a review article *Journal of Marketing Research* 5 (February), 83-98.

Frank, R.E., W.F. Massy and D. Morrison (1965), "Bias in multiple discriminant analysis", *Journal c Marketing Research* 2, 250-258.

Frank, R.E., W.F. Massy and Y. Wind (1972), *Market segmentation*, Prentice Hall, Englewood Cliffs New Jersey.

Frank, R.E. and C.E. Strain (1972), "A segmentation research design using consumer panel data *Journal of Marketing Research* 9 (November), 285-390.

Fraser, C. and J.W. Bradford (1983), "Competitive market structure analysis: principal partitioning revealed substitutabilities", *Journal of Consumer Research* 10 (June), 15-30.

Funkhouser, G.R. (1983), "A note on the reliability of certain clustering algorithms", *Journal of Marketing Research* 20 (February), 99-102.

Gensch, D.H. (1978), "Image-measurement segmentation", *Journal of Marketing Research* 15 (August), 384-394.

Gensch D.H. (1985), "Empirically testing a disaggregate choice model for segments", *Journal of Marketing Research* 22 (November), 462-467.

Gentry, J.W. and A.C. Burns (1977), "How important are evaluative criteria in shopping center patronage?", *Journal of Retailing* 53 (4), 73-86, 94-95.

Ginter, J.L. and E.A. Pessemier (1978), "Brand preference segments", *Journal of Business Research* 6, 111-131.

Goldman, A. (1977), "The shopping style explanation for store loyalty", *Journal of Retailing* 53 (4), 33-46, 94.

Goodman, L.A. (1974), "Exploratory latent structure analysis using both identifiable and unidentifiable models", *Biometrika* 61 (2), 215-231.

Gordon A.D. (1987), "A review of hierarchical classification", *Journal of Royal Statistical Soc. A* 150, 119-138.

Green, P.E. (1977), "A new approach to market segmentation", *Business Horizons* 20, 61-73.

Green, P.E. and F.J. Carmone (1969), "Multidimensional scaling: an introduction and comparison of nonmetric unfolding techniques", *Journal of Marketing Research* 6 (August), 330-341.

Green, P.E. and F.J. Carmone (1977), "Segment congruence analysis: a method for analyzing association among alternative bases for market segmentation", *Journal of Consumer Research* 3 (March), 217-222.

Green, P.E., F.J. Carmone and D.P. Wachspress (1976), "Consumer segmentation via latent class analysis", *Journal of Consumer Research* 3 (December), 170-174.

Green, P.E. and W.S. DeSarbo (1979), "Componential segmentation in the analysis of consumer trade-offs", *Journal of Marketing* 43, 83-91.

Green, P.E. and V. Srinivasan (1978), "Conjoint analysis in consumer research: issues and outlook", *Journal of Consumer Research* 5, 103-123.

Green, P.E. and Y. Wind (1973), *Multiattribute decisions in marketing: a measurement approach*, Dryden Press, Hinsdale, Illinois.

Green, P.E., Y. Wind and H.J. Claycamp (1975), "Brand features congruence mapping", *Journal of Marketing Research* 12 (August), 306-313.

Green, P.E., Y. Wind, A.K. Jain (1972), "Benefit bundle analysis", *Journal of Advertising Research* 12, 31-36.

Grover, R. and V.R. Rao (1988), "Inferring competitive market structure based on a model of interpurchase intervals", *International Journal of Research in Marketing* 5, 55-72.

Grover, R. and V. Srinivasan (1987), "A simultaneous approach to market segmentation and market structuring", *Journal of Marketing Research* 24, 139-153.

Grover, R. and V. Srinivasan (1989), "An approach for tracking within-segment shifts in market shares", *Journal of Marketing Research* 26 (May), 230-236.

Gunderson, R.W. (1982), "Choosing the r-dimension for the FCV family of clustering algorithms", *BIT* 22, 140-149.

Hagerty M.R. (1985), "Improving the predictive power of conjoint analysis: the use of factor analysis and cluster analysis", *Journal of Marketing Research* 22 (May), 168-184.

Haley, R.I. (1984), "Benefit segments: backwards and forwards", *Journal of Advertising Research* 24 (1), 19-25.

Haley, R.I. (1968), "Benefit segmentation: a decision-oriented research tool", *Journal of Marketing* 32 (July), 30-35.

Hansen, R.A. and T. Deutscher (1977), "An empirical investigation of attribute importance in retail store selection", *Journal of Retailing* 53 (4), 59-72.

Hartigan, J.A. (1975), *Clustering Algorithms*, Wiley, New York.

Hauser, J.R. and F.S. Koppelman (1979), "Alternative perceptual mapping techniques: relative accuracy and usefulness", *Journal of Marketing Research* 16, 495-506.

Hauser, J.R. and G.L. Urban (1977), "A normative methodology for modeling consumer response to innovation", *Operations Research* 25 (4), 579-617.

Herraman, C. (1968), "Normalizing a symmetric matrix" (Algorithm AS11), *Applied Statistics* 17, 287-288.

Holbrook, M.B. and D.V. Holloway (1984), "Marketing strategy and the structure of aggregate, segment-specific, and differential preferences", *Journal of Marketing* 48, 62-67.

Holbrook, M.B. and J.M. Hubert (1975), "Multi-attribute models: a comparative analysis", *Consumer Research* 2, 375-388.

Hope, A.C.A. (1968), "A simplified Monte Carlo significance test procedure", *Journal of the Royal Statistical Society, Series B* 30, 582-598.

Howard, J.A. (1985), *Consumer behavior in marketing strategy*, Prentice-Hall International, Inc., New York.

Hruschka, H. (1986), "Market definition and segmentation using fuzzy clustering methods", *International Journal of Research in Marketing* 3, 117-134.

Hustad, T.P., C.S. Mayer and T.W. Whipple (1975), "Consideration of context differences in product evaluation and market segmentation", *Journal of the Academy of Marketing Science* 3 (1), 34-47.

Inglis, J. and D. Johnson (1970), "Some observations on and developments in the analysis of multisurvey data", *Journal of the Market Reseach Society* 12, 75-88.

James, D.L., R.M. Durand and R.E. Dreves (1976), "The use of a multi-attribute model in a store image study", *Journal of Retailing* 52 (2), 23-32.

Johnson, R.M. (1971), "Market segmentation: a strategic management tool", *Journal of Marketing Research* 8 (February), 13-18.

Kamakura, W.A. (1988), "A least squares procedure for benefit segmentation with conjoint experiments", *Journal of Marketing Research* 25, 157-167.

Kamakura, W.A. and J. Agrawal (1990), "A clusterwise multinomial logit model for benefit segmentation", *Working Paper, Van der Bilt University*, Nashville.

Kamakura, W.A. and G.J. Russell (1989), "A probabilistic choice model for market segmentation and elasticity structure", *Journal of Marketing Research* 26 (November), 379-390.

Kapferer, J. and G. Laurent (1985), "Consumer involvement profiles: a new practical approach to consumer involvement", *Journal of Advertising Research* 25 (6), 48-56.

Kistemaker, C. and M. Wedel (1988), *RMSCLUST users manual.* Report number V 88.316 of the TNO-CIVO Institutes, Zeist, The Netherlands.

Klahr, D. (1969), "A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure", *Psychometrika* 34 (3), 319-330.

Koponen, A. (1960), "Personality characteristics of purchasers", *Journal of Advertising Research* 1, 6-12.

Kotler, P. (1988), *Marketing Management*, Prentice-Hall, Englewood Cliffs, New Jersey.

Kotrba, R.W. (1966), "The strategy selection chart", *Journal of Marketing* 30 (July), 89-95.

Kruskal, J.B. (1964), "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis", *Psychometrika* 29 (1), 1-27.

Lastovicka J.L. (1982), "On the validation of lifestyle traits: a review and illustration", *Journal of Marketing Research* 19, 126-138.

Lastovicka J.L., J.P. Murry jr. and E.A. Joachimsthaler (1990), "Evaluating the measurement validity of lifestyle typologies with qualitative measures and multiplicative factoring", *Journal of Marketing Research* 27 (February), 11-23.

Lazer W. (1963), "Lifestyle concepts and marketing", in: S. Greyser (ed.), *Toward Scientific Marketing*, American Marketing Association, Chicago, 130.

Leeflang P.S.H. (1974), *Mathematical models in marketing*, Stenfert Kroese B.V., Leiden, Holland.

Leeflang P.S.H. (1987), *Probleemgebied marketing, een management benadering*, Stenfert Kroese B.V., Leiden, Holland.

Lessig V.P. and J.O. Tollefson (1971), "Market segmentation through numerical taxonomy", *Journal of Marketing Research* 8 (November), 480-487.

Lilien, G.L. and P. Kotler (1983), *Marketing decision making*, Harper and Row, New York.

Loudon, D. and A.J. Della Bitta (1984), *Consumer Behavior. Concepts and applications*, McGraw-Hill International Editions, London.

Lutz, J.G. (1977), "The multivariate analogue of JAN", *Educational and Psychological Measurement* 37 (1), 37-45.

Lynch J.G., Jr. (1985), "Uniqueness issues in the decompositional modeling of multiattribute overall evaluations: an information integration perspective", *Journal of Marketing Research* 22 (February), 1-19.

MacLachlan, D.L. and J.K. Johansson (1981), "Market segmentation with multivariate AID", *Journal of Marketing* 45, 74-84.

Mahajan, V. and A.K. Jain (1978), "An approach to normative segmentation", *Journal of Marketing Research* 15 (August), 338-345.

Mardia, K.V., J.T. Kent and J.M. Bibby (1988), *Multivariate analysis*, Academic Press United, London.

Martens, M. and H. Martens (1986), "Partial least squares regression", in: J.R. Piggott (ed.), *Statistical Procedures in Food Research*, Elsevier Applied Science, London.

Martin, C.R. and R.L. Wright (1974), "Profit-oriented data analysis for market segmentation: an alternative to AID", *Journal of Marketing Research* 11 (August), 237-242.

Martineau, P. (1958), "The personality of a retail store", *Harvard Business Review* 36 (January-February), 47-55.

Mason, R.L., R.F. Gunst and J.T. Webster (1975), "Regression analysis and problems c multicollinearity", *Communications in Statistics* 4 (3), 277-292.

Massy, W.F. and R.E. Frank (1965), "Short term price and dealing effects in selected market segments' *Journal of Marketing Research* 2 (May), 171-185.

Massy, W.F., D.B. Montgomery and D.G. Morrison (1970), *Stochastic models of buying behavior*, The MIT Press, Cambridge, Massachusetts.

McCann, J.M. (1974), "Market segment response to the marketing decision variables", *Journal o Marketing Research* 11 (November), 399-412.

McCullagh, P. and J.A. Nelder (1989), *Generalized Linear Models*, Chapman and Hall, New York.

McElwain, D.W. and J.A. Keats (1961), "Multidimensional unfolding: some geometrical solutions' *Psychometrika* 26 (3), 325-332.

McLachlan, G.J. and K.E. Basford (1988), *Mixture Models: Inference and Applications to Clustering*, Marcel Dekker, Inc., New York.

Miller, K.E. and J.L. Ginter (1979), "An investigation of situational variation in brand choice behavic and attitude', *Journal of Marketing Research* 16 (February), 111-123.

Milligan, G.W. and M.C. Cooper (1985), "An examination of procedures for determining the number c clusters in a dataset", *Psychometrika* 50 (2), 159-179.

Mitchel, A. (1983), *The nine American lifestyles*, New York, MacMillan Co.

Moore, W.L. (1980), "Levels of aggregation in conjoint analysis: an empirical comparison" *Journal o Marketing Research* 17 (November), 516-523.

Moriarty, M. and M. Venkatesan (1978), "Concept evaluation & market segmentation", *Journal o Marketing* 42 (July), 82-86.

Morrison, D.G. (1973), "Evaluation market segmentation studies: the properties of R^2", *Managemen Sciences* 19 (11), 1213-1221.

Neslin, S.A. (1981), "Linking product features to perceptions: Self-stated versus statistically reveale importance weights", *Journal of Marketing Research* 18 (February), 80-86.

Nunnally, J.C. (1967), *Psychometric Theory*, New York McGraw-Hill.

Ogawa, K. (1987), "An approach to simultaneous estimation and segmentation in conjoint analysis' *Marketing Science* 6 (1), 66-81.

Oliva, T.A. and R.E. Reidenbach (1985), "Iterative partitioning methods: the use of mapping theory as clustering technique", *Journal of Marketing Research* 22 (February), 81-85.

Oppedijk van Veen, W.M., and T.M.M. Verhallen (1986), "Vacation market segmentation: a domair specific value approach", *Annuals of Tourism Research* 13, 37-58.

Pedhazur, E.J. (1982), *Multiple regression in behavioral research*, Holt, Rinehart and Winston, New York.

Pieters, R.G.M. (1989), *Attitudes and behavior in a source-separation program,: a garbology approach*, Eburon, Delft.

Plat, F.W. and P.S.H. Leeflang (1988), "Decomposing sales elasticities on segmented markets *International Journal of Research in Marketing* 5 (April), 303-315.

Plummer, J.T. (1974), "The concept and application of life style segmentation", *Journal of Marketing* 3E (January), 33-37.

Punj, G. and D.W. Stewart (1983), "Cluster analysis in marketing research: review and suggestions for application", *Journal of Marketing Research* 20 (May), 134-148.

Rao, V.R. and F.W. Winter (1978), "An application of the multivariate probit model to market segmentation and product design", *Journal of Marketing Research* 15 (August), 361-368.

Richardson, D.P. (1987), "Effects of nutritional guidelines on food marketing and new product development". In: Solms, S. et al. (eds.), *Food acceptance and nutrition*, Academic press, London.

Robinson J. (1938), *The Economics of Imperfect Competition*, McMillan and Co. Limited, London.

Rosenberg, M.J. (1960), "A structural theory of attitude dynamics", *Public Opinion Quarterly* 24, 319-340.

Rotter, J.B. (1966), "Generalized expectancy for internal versus external control of reinforcement", *Psychological Monographs* 80, 1-27.

Rotter, J.B., J.E. Chance and E.J. Phares (1972), *Applications of a Social Learning Theory of Personality*, Holt, Rinehart, and Winston, New York.

Ruspini, E. (1969), "A new approach to clustering", *Inf. and Conf.* 15, 22-32.

Sawtooth Software (1986), Ci2 System: Version 2, Ketchum (Idaho).

Schervish, M.J. (1987), "A review of multivariate analysis", *Statistical Science* 2 (4), 396-433.

Schiffman, L.G., J.F. Dash and W.R. Dillon (1977), "The contribution of store-image characteristics to store-type choice", *Journal of Retailing* 53 (2), 3-14, 46.

Scott, A.J. and M.J. Symons (1971), "Clustering methods based on likelihood ratio criteria", *Biometrics* 27, 387-397.

Sewall, M.A. (1978), "Market segmentation based on consumer ratings of proposed product designs", *Journal of Marketing Research* 15, (November), 557-564.

Sexton, D.E., Jr. (1974), "A cluster analytic approach to market response function", *Journal of Marketing Research* 11 (February), 109-114.

Shepard, R.N. and P. Arabie (1979), "Additive clustering: representation of similarities as combinations of discrete overlapping properties", *Psychological Review* 86 (2), 87-123.

Sheth, J.N. (1974), *Multivariate Methods for Market and Survey Research*, American Marketing Association, University of Illinois, Chicago.

Shocker, A.D. and V. Srinivasan (1974), "A consumer-based methodology for the identification of new product ideas", *Management Science* 20 (6), 921-937.

Silk, A.J. and G.L. Urban (1978), "Pretest market evaluation of new packaged goods: a model and measurement methodology", *Journal of Marketing Research* 15 (May), 171-191.

Slovic, P., and S. Lichtenstein (1971), "Comparison of Bayesian and regression approaches to the study of information processing in judgment", *Organizational Behavior and Human Performance* 6, 649-744.

Smith, W.R. (1956), "Product differentiation and market segmentation as alternative marketing strategies", *Journal of Marketing* 21 (July), 3-8.

Späth, H. (1977), "Computational experiences with the exchange method", *European Journal of Operational Research* 1, 23-31.

Späth, H. (1979), "Clusterwise linear regression", *Computing* 22 (4), 367-373.

Späth, H. (1981), "Clusterwise linear regression", *Computing* 26 (26), 275.

Späth, H. (1982), "A fast algorithm for clusterwise linear regression", *Computing* 29 (2), 175-181.

Späth, H. (1986), "Clusterwise linear least absolute deviations regression", *Computing* 37, 371-378.

Srivastava, R.K., M.I. Alpert and A.D. Shocker (1984), "A customer-oriented approach for determinin market structures", *Journal of Marketing* 48, 32-45.

Starr, M.K. and J.R. Rubinson (1978), "A loyalty group segmentation model for brand purchasin simulation", *Journal of Marketing Research* 15 (August), 378-383.

Steenkamp, J.B.E.M. (1987), "Perceived quality of food products and its relationship to consume preferences: theory and measurement", *Journal of Food Quality* 9, 373-386.

Steenkamp, J.B.E.M. (1989), *Product Quality*, Van Gorcum, Assen.

Steenkamp, J.B.E.M. (1990), "A conceptual model of the quality perception process", *Journal o Business Research* (in press).

Stewart, D. W. (1981), "The application and misapplication of factor analysis in marketing research" *Journal of Marketing Research* 17, February, 51-62.

Stout, R.G. et al (1977), "Usage incidence as a basis for segmentation", in: Wind Y. and Greenberg, M (eds), *Moving Ahead with Attitude Research*, American Marketing Association, Chicago, 45-49.

Takeuchi, H. and J.A. Quelch (1983), "Quality is more than making a good product", *Harvard Business Review* 61 (July-August), 139-145.

Talarzyk, W.W. (1972), "A reply to the response to Bass, Talarzyk, and Sheth", *Journal of Marketing Research* 9, 465-467.

Tantiwong, D. and P.C. Wilton (1985), "Understanding food preferences among the elderly using hybri conjoint measurement models", *Journal of Retailing* 61 (4), 35-64.

Theil, H. (1971), *Principles of econometrics*, Wiley, New York.

Thrane, K.E. and R.W. Gunderson (1986), "Source allocation of organic air pollutants by application o fuzzy c-varieties pattern recognition", *Analytica Chimica Acta* 191, 307-317.

Tollefson J.O. and V.P. Lessig (1978), "Aggregation criteria in normative market segmentation theory" *Journal of Marketing Research* 15 (August), 346-355.

Twedt, D.W. (1967), "How does awareness-attitude affect marketing strategy?", *Journal of Marketing* 31, (October), 64-66.

Tybout, A.M. and J.R. Hauser (1981), "A marketing audit using a conceptual model of consumer be havior: application and evaluation", *Journal of Marketing* 45, 82-101.

Umesh, U.N. (1987), "Transferability of preference models across segments and geographic areas" *Journal of Marketing* 51, 59-70.

Urban, G.L. (1975), "PERCEPTOR: A model for product positioning", *Management Science* 21 (April), 858-871.

Urban, G.L. and J.R. Hauser (1980), *Design and Marketing of New Products,* Prentice Hall, Englewood Cliffs, N.J.

Urban, G.L., J.R. Hauser and N. Dholakia (1987), *Essentials of new product management*, Prentice Hall, Englewood Cliffs, N.J.

Verhallen, T.M.M. and G.J. DeNooy (1982), "Retail attribute sensitivity and shopping patronage" *Journal of Economic Psychology* 2, 39-55.

Verhallen, T.M.M. and R.G.M. Pieters (1984), "Attitude theory and behavioral costs", *Journal o Economic Psychology* 5, 223-249.

WBO (1985), Demo 84/85: *Bevolkingsgegevens en Normvraagstellingen voor Markt- en Opinie-Onderzoek.*

Wedel, M. and C. Kistemaker (1989), "Consumer benefit segmentation using clusterwise linear regression", *International Journal of Research in Marketing* 6, 45-59.

Wedel, M., K.F.A.M. Hulshof and M.R.M. Löwik (1986), "Determinants of choice behavior of elderly people in the Netherlands with respect to meat products", in: *Contemporary research in marketing, proceedings of the XV-th annual conference of the European Marketing Academy*: Helsinki, 603-605.

Wedel, M., and J.B.E.M. Steenkamp (1989), "Fuzzy clusterwise regression approach to benefit segmentation", *International Journal of Research in Marketing* 6, 241-258.

Wells, W.D. (1975), "Psychographics: a critical review", *Journal of Marketing Research* 12 (May), 196-213.

Wierenga, B. (1983), "Model and measurement methodology for the analysis of consumer choice of food products", *Journal of Food Quality* 6, 119-137.

Wildt A.R. (1976), "On evaluating market segmentation studies and the properties of R^2", *Management Science* 22, (April), 904-908.

Wildt A.R. and and J.M. McCann (1980), "A regression model for market segmentation studies", *Journal of Marketing Research* 17 (August), 335-340.

Wilkie, W.L. (1970), *An empirical analysis of alternative basis for market segmentation*, unpublished doctoral dissertation, Stanford University.

Wilkie, W.L. and Cohen J.B. (1977), *An overview of market segmentation: behavioral concepts and research approaches*, Working Paper, Marketing Science Institute, Cambridge, Massachusetts.

Wilkie, W.L. and E.A. Pessemier (1973), "Issues in marketing's use of multi-attribute attitude models", *Journal of Marketing Research* 10 (November), 428-441.

Wind, Y. (1978), "Issues and advances in segmentation research", *Journal of Marketing Research* 15 (August), 317-337.

Winter, F.W. (1979), "A cost-benefit approach to market segmentation", *Journal of Marketing* 43, 103-111.

Winter, F.W. (1984), "Market segmentation: a tactical approach", *Business Horizons* 27 (January-February), 57-63.

Wold, H. (1982), "Soft modelling: the basic design and some extensions", in: Karl G. Jöreskog and Herman Wold, (eds.), *System under indirect observation* Vol. II, North-Holland, Amsterdam, pp. 1-54.

Wolfe, J.H. (1970), "Pattern clustering by multivariate mixture analysis", *Multivariate Behavioral Research* 5 (July), 329-350.

Yankelovich, D. (1964), "New criteria for market segmentation", *Harvard Business Review* 42, 83-90.

Young, S., L. Ott B. and Feigin (1978), "Some practical considerations in market segmentation", *Journal of Marketing Research* 15 (August), 405-412.

Zadeh, L.A. (1965), "Fuzzy sets", *Information and Control* 8, 338-353.

Zaichkowsky, J.L. (1985), "Measuring the involvement construct", *Journal of Consumer Research* 12 (December), 341-352.

Zaichkowsky, J.L. (1987), "The personal involvement inventory: reduction, revision, and application to advertising, *Working Paper, Simon Fraser University*, Burnaby, Canada.

Ziff, R. (1971), "Psychographics for market segmentation", *Journal of Advertising Research* 11 (2), 3-9.

SUMMARY

The present work consists of two major parts. In the first part the literature on market segmentation is reviewed, in the second part a set of new methods for market segmentation are developed and applied.

Part 1 starts with a discussion of the segmentation concept, and proceeds with a discussion on marketing strategies for segmented markets. A number of criteria for effective segmentation are summarized. Next, two major streams of segmentation research are identified on the basis of their theoretical foundation, which is either of a microeconomic or of a behavioral science nature. These two streams differ according to both the bases and the methods used for segmenting markets.

After a discussion of the segmentation bases that have been put forward as the normative ideal but have been applied in practice very little, different bases are classified into four categories, according to their being observable or unobservable, and general or product-specific. The bases in each of the four categories are reviewed and discussed in terms of the criteria for effective segmentation. Product benefits are identified as one of the most effective bases by these criteria.

Subsequently, the statistical methods available for segmentation are discussed, according to a classification into four categories, being either a priori or post hoc, and either descriptive or predictive. Post hoc (clustering) methods are appealing because they deal adequately with the complexity of markets, while the predictive methods within this class (AID, clusterwise regression) combine this advantage with prediction of purchase (predisposition).

Within the two major segmentation streams, segmentation methods have been developed that are specifically tailored to the segmentation problems at hand. These are discussed. For the microeconomic school focus is upon recently developed latent class approaches that simultaneously estimate consumer segments and market characteristics (market shares, switching, elasticities) within these segments. For the behavioral science school focus is on benefit segmentation. Disadvantages of the traditional two-stage approach, in which consumers are clustered into segments on the basis of benefit importances estimated at the individual level, are revealed and procedures that have been addressed to one or more of these problems are reviewed.

In Part 2, three new methods for benefit segmentation are developed: clusterwise regression, fuzzy clusterwise regression and generalized fuzzy clusterwise regression.

The first method is a clustering method that simultaneously groups consumers in a number of nonoverlapping segments, and estimates the benefit importances within segments. The performance of the algorithm on synthetic data is investigated in a Monte Carlo study. Empirically, the method is shown to outperform the two-stage procedure. Special attention is paid to significance testing with Monte Carlo test procedures, and convergence to local optima. An application to segmentation of the meat-market in the Netherlands on the basis of data on elderly peoples preferences for meat products is given. Three segments are identified. The first segment weights sensory quality against exclusiveness (price), in the second segment quality is traded off against fatness. This segment, comprising predominantly of females, had the best knowledge of nutrition. In the third segment preference is based on quality only. Regional differences were identified among segments.

Fuzzy clusterwise regression extends clusterwise regression in that it allows consumers to be a member of more than one segment. It simultaneously estimates the preference functions within segments, as well as the degree of membership of consumers in those segments. Using synthetic data, the performance of the method is evaluated. Empirical comparisons with two other methods are provided, and the cross-validity of the method with respect to classification and prediction is assessed. Attention is given in particular to the selection of the appropriate number of segments, the setting of the user defined fuzzy weight parameter, and Monte Carlo significance test procedures. An application to data on preferences for meat-products used on bread in the Netherlands revealed three segments. In the first segment, taste and fitness for common use are important. In the second segment, taste overridingly determines preference, but products that are considered more exclusive and natural and less fat and salt are also preferred. In segment three the health related product benefits are even more important. The importance of taste decreases from segment one to three, while the importance of health-related aspects increases in that direction. The health oriented segments comprised more females, older people and people who attributed causality of their behavior more to themselves.

The method was also applied to data on consumers image for stores that sell meat. Again three segments were revealed. The value shoppers, trade off quality and price.

They come from smaller families and spend less on meat. In the largest segment store image is based upon product quality. Females have higher membership in this segment, that is more involved with the store where they buy meat. For service shoppers, both service and atmosphere are important. This segment tends to be more store-loyal.

Next, a generalization of fuzzy clusterwise regression is proposed, which incorporates both benefit segmentation and market structuring within the framework of preference analysis. The method simultaneously estimates the preference functions within each of a number of clusters, and the parameters indicating the degree of membership of both subjects and products in these clusters. The performance of this method is assessed in a Monte Carlo study on synthetic data. The method is compared empirically with clusterwise regression and fuzzy clusterwise regression. The significance testing with Monte Carlo test procedures, and the selection of the fuzzy weight parameters is treated in detail. Two segments were revealed in an analysis of consumer preferences of butter and margarine brands. The segments differed mainly in the importance attached to exclusiveness and fitness for multiple purposes. The brands competing within these segments were revealed. Females and consumers with a higher socioeconomic status had higher memberships in the segments in which exclusiveness was important.

Finally, the clusterwise regression methods developed in this work are compared with other recently developed procedures in terms of the assumptions involved. The substantive results obtained in the empirical studies concerning foods are summarized and their implications for future research are given. The implications and the contribution of the methods to the development of marketing strategies for segmented markets are discussed.

SAMENVATTING

Dit werk bestaat uit twee delen. In het eerste deel wordt een overzicht gegeven van de literatuur over marktsegmentatie, in het tweede deel wordt een set nieuwe methoden voor marktsegmenten ontwikkeld en toegepast.

Deel 1 begint met een beschouwing over het segmentatieconcept, waarna een beschrijving wordt geven van marketingstrategieën voor gesegmenteerde markten. Een aantal criteria voor effectieve segmentatie worden samengevat.

Vervolgens worden twee hoofdstromen in het segmentatieonderzoek onderscheiden op basis van hun theoretische achtergrond, die micro-economisch of gedragswetenschappelijk kan zijn. Deze twee stromen verschillen zowel in de bases als in de methoden die worden gebruikt voor het segmenteren van markten.

Na een bespreking van de bases die gepostuleerd zijn als normatief ideaal, maar die in de praktijk weinig zijn toegepast, worden een viertal categorieën van bases onderscheiden al naar gelang ze direct waarneembaar zijn en al naar gelang ze algemeen zijn of produkt-specifiek. Er wordt vervolgens een overzicht gegeven van de bases in elk van de vier categorieën, die worden besproken in termen van de criteria voor effectieve segmentatie. Het nut dat produkten hebben voor consumenten wordt geïdentificeerd als een van de meest effectieve bases.

Vervolgens worden de statistische methoden die voor segmentatie gebruikt kunnen worden, besproken aan de hand van een indeling in vier klassen, al naar gelang de bases a priori of post-hoc een segmentatie opleveren en al naar gelang ze beschrijvend of voorspellend zijn. De post-hoc (clustering)- methoden zijn aantrekkelijk omdat ze op adequate wijze de complexiteit van markten beschrijven, terwijl de voorspellende methoden binnen deze groep (AID, clusterwise regression) dit voordeel combineren met het voorspellen van aankoop-(intenties).

Binnen elk van de twee hoofdstromen van segmentatieonderzoek zijn methoden ontwikkeld die zijn toegesneden op specifieke problemen. Deze worden besproken. Binnen het micro-economisch georiënteerde segmentatieonderzoek wordt nadruk gelegd op de recent ontwikkelde latente-klassemodellen, die tegelijkertijd segmenten identificeren en grootheden zoals marktaandelen, merktrouw en elasticiteiten binnen deze segmenten schatten. Bij de bespreking van het gedragswetenschappelijk georiënteerde segmentatieonderzoek ligt de nadruk op segmentatie op basis van het nut van produkten. Nadelen van de traditionele twee-staps benadering, waarin consumenten in segmenten worden gegroepeerd op basis van het op individueel niveau geschatte belang van produktkenmerken, worden geïdentificeerd. Procedures die zich op één of meer van deze nadelen richten worden besproken.

In deel 2 worden drie nieuwe methoden voor nutssegmentatie ontwikkeld: "clusterwise regression", "fuzzy clusterwise regression", en "generalized fuzzy clusterwise regression".

De eerste methode is een clusteringmethode die tegelijkertijd consumenten in een aantal niet-overlappende segmenten groepeert en het belang van produkt-eigenschappen in elk van deze segmenten schat. De prestaties van het algoritme op synthetische data worden onderzocht in een Monte Carlo studie. In een empirische studie wordt gedemonstreerd dat de methode zich gunstig verhoudt ten opzichte van de twee-staps procedure. Speciale aandacht wordt besteed aan significante toetsen met Monte Carlo testprocedures en convergentie naar lokale optima. De methode wordt toegepast om van de vleesmarkt in Nederland de segmenten op basis van gegevens over de voorkeur van oudere mensen voor vlees te beschrijven. Drie segmenten worden geïdentificeerd. Het eerste segment weegt sensorische kwaliteit af tegen exclusiviteit (prijs), in het tweede segment wordt kwaliteit af-gewogen tegen vetheid. Dit segment, dat voornamelijk uit vrouwen bestaat, heeft ook de meeste kennis van voeding. In het derde segment is de voorkeur uitsluitend gebaseerd op sensorische kwaliteit. Regionale verschillen tussen segmenten worden geïdentificeerd.

Fuzzy clusterwise regression is een uitbreiding van clusterwise regression daar deze methode toestaat dat consumenten tot meer dan één segment behoren. De methode schat tegelijkertijd de preferentiefuncties in een aantal segmenten en de mate van lidmaatschap van consumenten in deze segmenten.

De prestaties van de methode worden onderzocht met synthetische data. De methode wordt empirisch vergeleken met twee andere methoden en de kruis-validiteit met betrekking tot classificatie en predictie worden onderzocht. Er wordt extra aandacht besteed aan de problematiek met betrekking tot de selectie van het aantal segmenten, de "fuzzy weight" parameters die door de gebruiker moet worden ingesteld, en Monte Carlo testprocedures.

Een toepassing op gegevens over voorkeur voor vleeswaren in Nederland onthulde drie segmenten. In het eerste segment zijn smaak en geschiktheid voor dagelijks gebruik belangrijk. In het tweede segment bepaalt smaak in belangrijke mate de voorkeur, maar produkten die als exclusiever, natuurlijker en minder vet en zout worden beschouwd hebben ook een hogere voorkeur. In het derde segment zijn deze gezondheidsgerelateerde aspecten nog belangrijker. Het belang van smaak neemt af van segment één naar drie, terwijl het belang van gezondheidsaspecten in die richting toeneemt. De gezondheidsgeoriënteerde segmenten bevatten meer vrouwen, ouderen en mensen die de consequenties van hun gedrag meer aan zichzelf toeschrijven dan aan anderen.

De methoden wordt eveneens toegepast op gegevens over het imago van winkels die vlees verkopen. Opnieuw worden drie segmenten geïdentificeerd. Het eerste segment weegt kwaliteit en prijs af. Deze consumenten komen uit kleinere gezinnen en besteden minder geld aan vlees en vleeswaren. In het tweede en grootste segment is het winkelimago gebaseerd op de kwaliteit van de produkten die verkocht worden. Vrouwen hebben een hoger lidmaatschap in dit segment, dat meer betrokken is bij de winkel waar het vlees gekocht wordt. In het derde segment zijn zowel service als atmosfeer belangrijk. Dit segment vertoont een grotere mate van winkeltrouw.

Vervolgens wordt een generalisatie van fuzzy clusterwise regression voorgesteld, die zowel segmentatie als marktstructurering in preferentie-analyse incorporeert. De methode schat tegelijkertijd de preferentiefuncties in een aantal clusters, en de parameters die de mate van lidmaatschap van zowel personen als produkten in deze clusters aangeven. De prestaties van deze methode worden vastgesteld in een Monte Carlo studie op basis van synthetische data. De methode wordt empirisch vergeleken met clusterwise regression en fuzzy clusterwise regression.

De significante toetsen met Monte Carlo procedures en de selectie van de "fuzzy weight"-parameters wordt in detail behandeld. In een analyse van de voorkeur van consumenten voor boter en margarinemerken worden twee segmenten geïdentificeerd. De segmenten verschillen voornamelijk in het belang dat wordt gehecht aan exclusiviteit en geschiktheid voor dagelijks gebruik. De merken die binnen deze segmenten concurreren worden onthuld. Vrouwen en consumenten met een hogere sociaal-economische status hebben een hoger lidmaatschap in de segmenten waarin exclusiviteit belangrijk is.

Tenslotte worden de methoden die hier zijn ontwikkeld vergeleken met andere recent ontwikkelde methoden, met betrekking tot de aannamen. De resultaten die verkregen zijn in de empirische studies over voedingsmiddelen worden samengevat en hun implicaties voor toekomstig onderzoek worden gegeven. De implicaties en de bijdrage van de ontwikkelde methoden voor de ontwikkeling van strategieën voor gesegmenteerde markten worden gegeven.

CURRICULUM VITAE

Michel Wedel was born in The Hague on May 31, 1957. After having completed Atheneum B at the Thorbecke Lyceum in his native town in 1975, he started a study of Biology at the Leiden University. Systematic zoology, business management and biomathematics were his specializations. After completing this study he married Hennie Jansens in 1981. Since 1974, Michel has practiced Karate championships and gained a number of national and international titles. After retiring as an active sportsman in 1987, he accepted a position on the board and became the coach of the national team.

In 1982, Michel Wedel was appointed as a methodologist at the TNO-CIVO Toxicology and Nutrition Institute in Zeist. He was assigned the tasks of designing and analyzing studies on nutrition and health executed within the Department of Human Nutrition and, after a few years, giving guidance to people working in statistics and computer programming. In 1986, he became a statistician, after a four-year evening course under the auspices of the Netherlands Statistical Society.

In 1984, he was asked to develop models to analyze determinants of food purchase and consumption behavior, and was appointed as the manager of the project in question. Part of this work is reflected in this thesis. Since then his research interests are dedicated to projects on food marketing and nutrition information within the Department of Human Nutrition.