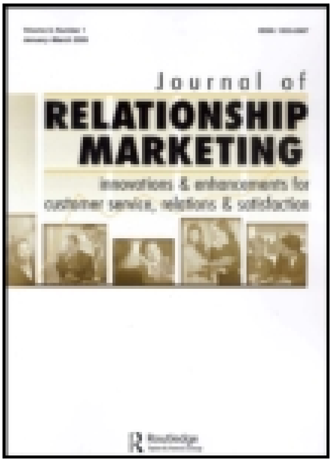


Journal of Relationship Marketing
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wjrm20

Approaches to Customer Segmentation

Bruce Cooil [a] , Lerzan Aksoy [b] & Timothy L. Keiningham [c]

[a] Owen Graduate School of Management, Vanderbilt University , Nashville, TN, 37203, USA

[b] Koç University, College of Administrative Sciences and Economics , Rumeli Feneri Yolu, Sariyer, 80910, Istanbul, Turkey

[c] IPSOS Loyalty, Morris Corporate Center 2 , 1 Upper Pond Road, Building D, Parsippany, NJ, 07054, USA

Published online: 25 Sep 2008.

Approaches to Customer Segmentation

Bruce Cooil

Vanderbilt University

Lerzan Aksoy

Koç University, Istanbul, Turkey

Timothy L. Keiningham

IPSOS Loyalty

SUMMARY. Customer segmentation has virtually unlimited potential as a tool that can guide firms toward more effective ways to market products and develop new ones. As a conceptual introduction to this topic, we study how an innovative multi-national firm (Migros Turk) has developed an effective set of segmentation strategies. This illustrates how

Bruce Cooil is Professor of Management, Owen Graduate School of Management, Vanderbilt University, Nashville, TN 37203 (E-mail: bruce.cooil@owen.vanderbilt.edu).

Lerzan Aksoy is Assistant Professor of Marketing, Koç University, College of Administrative Sciences and Economics, Rumeli Feneri Yolu, Sariyer 80910 Istanbul, Turkey (E-mail: laksoy@ku.edu.tr).

Timothy L. Keiningham is Senior Vice President & Head of Consulting, IPSOS Loyalty, Morris Corporate Center 2, 1 Upper Pond Road, Building D, Parsippany, NJ 07054 (E-mail: tim.keiningham@ipsos-na.com).

Bruce Cooil acknowledges support from the Dean's Fund for Faculty Research, Owen Graduate School, Vanderbilt University.

[Haworth co-indexing entry note]: "Approaches to Customer Segmentation." Cooil, Bruce, Lerzan Aksoy, and Timothy L. Keiningham. Co-published simultaneously in *Journal of Relationship Marketing* (Best Business Books, an imprint of The Haworth Press, Inc.) Vol. 6, No. 3/4, 2007, pp. 9-39; and: *Profit Maximization Through Customer Relationship Marketing: Measurement, Prediction and Implementation* (ed: Lerzan Aksoy, Timothy L. Keiningham, and David Bejou) Best Business Books, an imprint of The Haworth Press, Inc., 2007, pp. 9-39. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

Available online at http://jrm.haworthpress.com

doi:10.1300/J366v06n03_02

firms can construct novel and inventive approaches that provide great value. A-priori, and custom designed post-hoc methods are among the most important approaches that a firm should consider.

We then review general approaches to customer segmentation, with an emphasis on the most powerful and flexible analytical approaches and statistical models. This begins with a discussion of logistic regression for supervised classification, and general types of cluster analysis, both descriptive and predictive. Predictive clustering methods include cluster regression and CHAID (Chi-squared automatic interaction detection, which is also viewed as a tree classifier). Finally, we consider general latent class models that can handle multiple dependent measures of mixed type. These models can also accommodate samples that are drawn from a pre-defined group structure (e.g., multiple observations per household). To illustrate an application of these models, we study a large data set provided by an international specialty-goods retail chain.
doi:10.1300/J366v06n03_02 *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <http://www.HaworthPress.com>*

KEYWORDS. Latent class model, clustering, cluster regression, logistic regression, classification, conjoint analysis, random effect, multilevel model, inactive covariate, satisfaction

INTRODUCTION

Market segmentation can be defined as dividing a market into distinct groups of customers, with different needs, characteristics or behavior, who might require separate products or who may respond differently to various combinations of marketing efforts (Kotler & Armstrong, 1999). Some bases of segmentation that may be used include geographic, demographic, psychographic and behavioral. Other variables that may be used for segmentation include situational (e.g., purchase/use occasions), and customer preferences for products or specific product attribute levels. Effective segmentation usually requires that each segment be evaluated on certain criteria such as stability, growth potential, size, accessible, responsiveness, and whether the customers in that segment, and the marketing efforts directed toward them, are consistent with company objectives and resources (i.e., whether the segment is "actionable"). Segmentation is critical because a company has limited resources, and must focus on how to best identify and serve its customers. Individual customer segments are characterized by a certain degree of within-group homogeneity that helps ensure that the members of a segment will respond in similar ways to marketing efforts. This allows firms to more efficiently apply marketing resources to each segment. Of course, companies are motivated to undertake segmentation strategies only as long as these efforts provide a positive expected net payoff. In summary, effective segmentation allows a company to determine which customer groups they should try to serve and how to best position their products and services for each group. Consequently, segmentation is an integral part of the development of marketing objectives and strategies, where defining those objectives will generally include either (Ansoff, 1957; McDonald & Dunbar, 2004): (a) an analysis of how products should be sold or developed, based on an analysis of current customer segments, or (b) the identification of new segments as targets for existing products or for the development of new products.

Wedel and Kamakura (1998, Chapter 3) provide an extensive review of the literature on market segmentation, and carefully review each of several approaches, along with a discussion of the supporting statistical methodology. General approaches to segmentation include both a-priori and post-hoc methods.

1. A-priori segmentation methods require that segments be defined before data are collected. The segments may be determined using customer characteristics or product-specific information. Segments are then studied empirically using data that may provide additional customer information. In some cases, several alternative or overlapping segment bases, that were all defined a-priori, are compared and contrasted. The goal of such an analysis may be primarily descriptive (e.g., cross-tabulation, logistic regression), or it could include the development of models that use the predefined segments to predict one or more dependent variables.
2. Post-Hoc methods identify segments empirically through data analysis. Again the ultimate goal may be primarily to study the groups themselves, or it may be to develop a predictive model for a set of dependent variables.
3. There are also hybrid approaches that combine a-priori and post-hoc analyses (e.g., Green, 1977).

Objectives and Organization of the Following Sections

We will consider analytic approaches that can be used in each of these categories, but our emphasis will be on latent class models and other promising approaches for effective post-hoc descriptive and predictive analyses. We begin in the next section with conceptual examples of how one innovative firm has used customer segmentation. Then we begin our discussion of the analytic approaches to segmentation with a very brief summary of the most effective procedures for a-priori analyses. We refer to this as "segmentation based on supervised classification." In this framework, the a-priori definition of the segments provides a data set that is "supervised" in the sense that each customer is already classified into a segment, and the goal is to develop a model that allows one to classify new customers. This is followed by a brief review of how various types of cluster analysis have been used in post-hoc frameworks. We consider general clustering procedures that are not based on an explicit statistical model, which are among the primary methods used in post-hoc descriptive studies, but we also briefly summarize important predictive clustering approaches. The final section on methods will consider general latent class models that are appropriate for either descriptive or predictive post-hoc analyses, but which are especially flexible and powerful in post-hoc predictive studies. We conclude with two brief sections: a section summarizing how conjoint analysis provides a framework for segmentation analyses, and a summary discussion.

CONCEPTUAL EXAMPLES OF HOW AN INNOVATIVE FIRM USES CUSTOMER SEGMENTATION: MIGROS TURK T.A.Ş. IN TURKEY

A Brief History

Migros, currently the largest grocery chain in Turkey, was set up in 1954 via the joint initiatives of the Swiss Migros Cooperatives Union and Istanbul Municipality. Migros was founded for the mission of obtaining food supplies and consumables from producers under the supervision of the municipal authorities and to sell these products to inhabitants of Istanbul under hygienic conditions and at reasonable prices. In 1975, all of Migros shares were transferred to the Koç Group, one of the largest holding companies in Turkey.

Following this development, Migros engaged in a rapid expansionary strategy by increasing the number of stores in Istanbul, opening stores in other regions in Turkey and introducing a number of different store formats based on size and product variety. In addition, Migros introduced a number of stores under different brand names to cater to segments with a variety of needs. Şok discount stores were introduced in 1995 to expand the market and broaden the appeal to include the lower tier price-sensitive segment. In 1997, Migros also became one of the pioneers in cyber-shopping and introduced its virtual store, utilizing sophisticated infrastructure and technology.

In 2005, Migros merged with Tansaş, a successful local grocery chain, on the grounds that the combined company would be able to offer a value proposition to its customers at better terms and extract cost savings in sales, marketing and administrative functions through an efficient management strategy, given the challenging competitive market conditions of Turkey. By the end of 2006, Migros has 205 Migros, 228 Tansaş, 357 Şok, 8 Macrocenter stores in various service formats totaling 798 national stores.

In line with its customer-centric philosophy and Koç Group mission, the vision of Migros was defined to be *the closest to the customer* with various service formats and strategies designed to exceed customer expectations and maintain a market presence throughout Turkey and neighboring countries. To accomplish this objective and collect customer data, Migros initiated a loyalty program, the "Migros Club Card" by offering its customers the opportunity to become a member of a program where they accumulate points, redeem them for rewards, and benefit from periodic promotion and discount campaigns. Migros Club currently has about 4.5 million active members in its portfolio. Eighty percent of sales are generated through this loyalty program.

Segmentation Approaches

Migros employs a variety of different methods to effectively segment its customer base. These include value, behavioral, lifestyle, lifecycle and activity-based segmentation schemes. These approaches can undoubtedly be used individually, but usually multiple approaches are used in conjunction with each other. For any customer to be included in the pool of customers on which segmentation analyses are performed, he/she is required to have used a Migros Club card in the last three months prior to the analysis.

Approach 1: Value-Based Segmentation

This first approach uses the household as the unit of analysis. The base is divided into 9 profitability groups. Next, it's partitioned into 9 groups based on frequency of visit. These two measures are then used to construct a "productivity index." Finally, this productivity index is used, along with a measure of loyalty, to define six segments. Loyalty is defined in terms of whether the household has purchased from the basic food categories (such as fruit & vegetables, dairies, etc.) in the last 6 months. Figure 1 illustrates the three tiered, six-segment value pyramid created using the productivity and loyalty indices. Partitions 1, 3 and 5 are defined as loyal segments using the prior criteria. The most valuable customers in the pyramid (i.e., 1 and 2) typically constitute about 130,000 to 150,000 households. Some of the most popular applications of this segmentation approach include offering special treatment benefits to the most valued customers. The company has a reputation for offering innovative prizes such as movie gala tickets and opportunities to participate in wine tasting events specially organized for Migros customers. Based on this segmentation, Migros limits its win-back efforts to its best customers. Every month, former customers, who were previously in one of the most valuable segments, are sent an offer through Short Messaging Service (SMS).

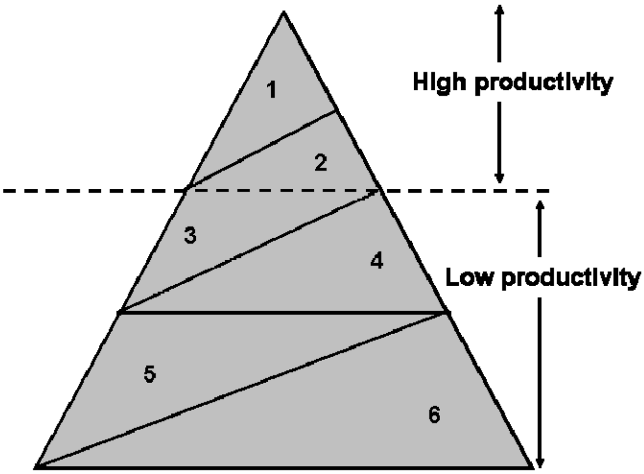
Approach 2: Behavioral-Based Segmentation

In contrast to the first approach, the behavioral based segmentation approach utilizes the individual as the unit of analysis. To be included in the base, it is necessary that customers have made a purchase, visited stores at least twice in the past 12 months. Cluster analyses techniques are utilized to determine groups of customers with similar behavioral patterns. Repeated quarterly iterations of the cluster analysis has revealed a consistent 7-cluster solution. Some of the variables typically included in the cluster analysis are listed below:

- Amount of Purchase
- Location and Purchase/Store Format
- Food versus non-food purchase
- Seasonality
- Degree of communication
- Products purchase

The basic differentiating characteristics of the clusters include but are not limited to amount spent, level of communication between the customer and staff, food versus non-food purchases and size of basket. For instance, clusters 3 and 6 are high-spending customers but differ in

FIGURE 1. The Value Pyramid



terms of the level of communication (limited versus extensive communication).

Approach 3: Lifestyle-Based Segmentation

This approach also uses the individual as the level of analysis. Contrary to the data driven approaches described earlier, the lifestyle segmentation scheme relies on an a-priori categorization of individuals and a profile of which products they are expected to buy. There are 14 predetermined lifestyle groups of which "gourmets" would be one such example. Gourmets are proposed to generally purchase high margin items such as cheese, fine wine and ethnic foods. Each individual is matched to the profile of one of the 14 segments and compared to a typical member of that segment based on the mean and standard deviation of purchase amounts, number of purchases, and the number of the predetermined items, as a proportion of total purchases. On the basis of this comparison, the individual is either included or considered a potential candidate to the segment. Hence, the number of resultant segments amounts to 28 where an actual and potential customer is ascribed to each of the 14 segments depending upon whether they meet the average and standard deviation criteria.

Although Migros has found 14 distinct lifestyle segments within its customer database, Migros strategically focuses on 4 segments, including "gourmets" and "diet lovers."

Approach 4: Activity Level-Based Segmentation

This analysis is performed on an individual level and utilizes a 15-month longitudinal data period to classify customers as either "active," "normal," or "passive" in terms of purchase amount and frequency. Initially, the first 12 months are used to calculate the average spending of the customer. Based on this average, the standard deviations are determined for the remaining months and for the latest 3 months. Then a weighted standard deviation is calculated, which gives more emphasis to changes in spending for more recent purchases. Those substantially lower than average are categorized as "passive" customers, those near the average are "normal" customer and those substantially higher than average are "active" customers. In addition to level of spending, the same procedure is repeated for frequency of purchase. Finally the two criteria are crossed to yield a 3 × 3 matrix that groups individuals into active, normal and passive groups in terms of both spending level and frequency. This segmentation is especially useful in conducting migration analysis. For example, special offers to increase their activity are made to customers who have migrated from the active to normal or normal to passive segments.

Success of Migros Turk

As a result of their effective segmentation scheme, Migros was not only able to improve its own productivity but also aide suppliers by providing certain analyses upon request, providing competitor comparisons and organizing joint campaigns. Migros also collaborates with partners such as banks, petrol stations for new customer acquisition efforts.

The high response rates to the campaigns organized by Migros, that are based on these segmentation methods, indicate that these are successful and viable approaches. A mailing campaign that is special to a unique segment can yield a response rate of 36%. In fact, the response rate can be increased up to 64% for some lifestyle segments such as "diet lovers." This rate is well above the average response rates of regular mail campaigns that are typically around 1-2%. Segmentation also helps Migros determine its strategic customer groups and activity plans.

Migros has been recognized both locally and internationally for its successful efforts. For example, in 2004, the company received an international award from "1 to 1 Impact" for its customer strategy and technology optimization.

REVIEW OF SEGMENTATION APPROACHES

Migros Turk is a great illustration of how a company can effectively employ novel segmentation approaches and obtain impressive business results. The next parts of this paper will discuss in more mathematical detail some of the most widely used and state-of-the-art methods in segmentation. After a brief discussion of the notation that will be used for customer variables, we study segmentation based on supervised classification, clustering methods, latent class models and conjoint analyses.

Notation for Customer Variables

In the following sections we will typically use terms like "customer attributes" or "customer variables," to refer to all possible types of information that can be used to identify segments. In expressions, we will represent these variables as a customer specific vector $\underline{z}$ or $\underline{z}_{i.}$. Also, in latent class models, where customer variables are also used as predictors for a dependent variable $\underline{y}_i$ (which may also be vector valued), we will generally distinguish between predictors and covariates as $\underline{z}_i^{(pred)}$ and $\underline{z}_i^{(cov)}$ respectively, where the covariates refer specifically to the set of customer variables that are used in the classification model that assigns customers to segments (or at least assigns segment membership probabilities to customers). In models where only predictors (e.g., cluster regression) or only covariates (e.g., supervised classification, or standard cluster analysis) are used, we will generally refer to these customer variables as simply "$\underline{z}$" without subscript. In all procedures, these variables, $\underline{z}$ (predictors or covariates), may include customer demographics, behavioral and psychographic information, socio-economic variables, product usage measures, brand loyalties, situational or transactional variables. They may also include attribute level preferences which may be defined a-priori or estimated as part of a conjoint analysis. Finally, $\underline{z}$ may include interactions among all of these types of variables.

SEGMENTATION BASED ON SUPERVISED CLASSIFICATION

Linear discriminant analysis, developed by Fisher (1936), may be the first statistical classification method in this category. Since then, many other methods have been developed, including the naïve Bayes' method, tree-based methods, neural networks, support vector machines, nonparametric kernel methods, and classification using ordinal or nominal logistic function analysis (Hastie, Tibshirani, & Friedman, 2001; Hand, 2006).

Certainly, one of the most flexible methods for K classes is the logistic model, where for each class k, we estimate a logistic regression. Let X_i represent the appropriate segment for a customer with attributes $\underline{z}_i$ (which may include situational or attribute level preference information), then the logit of $P[X_i = k]$ is assumed to be linear in subset of customer variables $\underline{z}_i^{(k)}$ that may be specific to segment k:

$$\text{Logit}\,([PX{=}k]/(1\text{-}P[X{=}k]){=}\ \underline{z}_i^{(k)}\underline{b}^{(k)} \tag{1}$$

which implies

$$P[X_i = k] = \exp\left(\underline{z}_i^{(k)}\underline{b}^{(k)}\right) / \sum_{\ell=1}^{K} \exp(\underline{z}_i^{(\ell)}\underline{b}^{(\ell)}) \tag{2}$$

One generally estimates these linear functions to maximize the multinomial likelihood, subject to the constraint that these probabilities are between 0 and 1, and $\sum_{\ell=1}^{K} P[X_i = k] = 1$, so it would suffice, for example, to constrain the K regression functions so that $\sum_{\ell=1}^{K} \underline{z}_i^{(\ell)}\underline{b}^{(\ell)} = 0$. In any case, classification of a customer to the appropriate segment is typically made on the basis of the largest estimated probability $P[X_i = k]$, but an advantage of logistic regression is that it also provides estimates of the membership probabilities themselves, as a function of customer attributes. In the section on latent class models, we will study approaches that are especially suited for ordinal classes (e.g., segments defined in terms of increasing profitability, or satisfaction), including the adjacent category model. (See equation (21) and the discussion that follows it.)

In the two-class case, Friedman, Hastie, and Tibshirani (2001) showed that the classification rule constructed from maximum likelihood estimation of the logistic regression model in (1)-(2) is equivalent to the best available boosting procedures. This is significant because boosting has been shown to provide very effective "enhanced" classification rules based on a weighted sum of weaker classification functions, each of which is based on the available set of covariates (customer variables). In the multiple-class case (K>2), classifiers based on maximizing the multinomial likelihood for logistic regression typically perform as well as the best multiclass boosting procedures (Friedman et al., 2001).

Logistic regression is generally considered to be a safer and more robust procedure than linear discriminant analysis (LDA) and its empirical performance is usually at least comparable (Hastie et al., 2001, pp. 103-105). Hand (2006) and Foody and Mathur (2004) compared the performance of LDA with other methods, on selected data sets. In two-class problems, Hand (2006) argued that LDA performs remarkably well relative to the best method for each data set. His relative performance measure was the proportional reduction in error (PRE) measure,

$$PRE = (E_0 - E_{LDA})/(E_0 - E_{Best})$$

where E_0, E_{LDA}, and E_{Best} are the misclassification rates for a default rule that assigns everything to the majority class, LDA, and the best known method, respectively. LDA's PRE was always at least 85% and usually above 90%. It was 100% for two of the 10 data sets. Nevertheless, the median and maximum values of E_{LDA}/ E_{Best} were 1.5 and 5.9 in this study, and as Friedman (2006) noted, even small reductions in error rate are sometimes very important in some applications. In Foody and Mathur (2004), training sets of different sizes are constructed to represented multiclass problems in remote-sensing (i.e., where information is gathered from aircraft, satellite, etc.). They showed that a neural network (NN) classifier and a multiclass support vector machine (SVM) classifier (that they develop) generally outperformed LDA for training sets of 15 to 100 observations (100 sets of each size). SVM usually slightly outperformed NN, although these differences were never statistically significant. Also, SVM's correct classification rate was generally about 2% higher than the LDA rate, irrespective of training set size. This difference was statistically significant ($p < 0.01$) only for classifications made when the training set size was 100. Zhong and Fukushima (2006) have also developed a new multiclass SVM method.

For a clear introduction to the SVM method, see Foody and Mathur (2004). SVM and NN classifiers share several disadvantages relative to logistic regression and LDA. These include the fact that they: (a) perform poorly at handling missing values, (b) do not easily accommodate data of mixed type, (c) are more sensitive to irrelevant covariates, and outliers, and (d) are harder to interpret (Hastie et al., 2001, pp. 312-314).

CLUSTERING METHODS

In this section, we consider methods of partitioning data into subgroups, that do not also provide a model for assigning observations to clusters. Clustering procedures that also provide a classification model will be considered later when we discuss general latent class models for segmentation.

Cluster analysis is one of the most popular methods for post-hoc descriptive studies and these methods use algorithms that can provide either overlapping or non-overlapping partitions, and among the non-overlapping methods both hierarchical and non-hierarchical approaches are possible. Non-hierarchical methods that use variance minimization as a way of defining groups, are among the most popular, and are frequently regarded as the most effective because they are more resistant to outliers and sample-specific anomalies (Punj & Stewart, 1983). Also, it is often difficult to motivate a hierarchical structure in terms of a business or marketing theory (Wedel & Kamakura, 1998).

Descriptive Clustering

The general post-hoc descriptive clustering approach is to partition the sample into a set S_K of K groups,

$$S_K \equiv \{G_1, \ldots, G_K\}, \tag{3}$$

that are selected to minimize some criterion $C(S_K)$. For example, one could choose the partition, S_K, that minimizes the weighted average within-group distance, or dissimilarity measure, D_{ij} between all pairs of observations i and j in each group:

$$C(S_K) = \sum_{k=1}^{K} w_k n_k^{-2} \sum_{i \in G_k} \sum_{j \in G_k} D_{ij}, \tag{4}$$

where w_k is group-specific weight which may itself simply be a function of the group size n_k. The D_{ij} may incorporate information from the multivariate attributes (x_{i1}, ..., x_{ip}) of each customer *i*. For example, following Friedman and Meulman (2004), one might define D_{ij} as a weighted average of attribute similarity measures like those developed by Gower (1971),

$$D_{ij} = \sum_{\ell=1}^{p} \omega_{\ell} d_{ijk}, \tag{5}$$

with

$$d_{ijk} \equiv \delta_{ijk} / s_k,$$

where typically

$$\begin{aligned} \delta_{ijk} &= |x_{ik} - x_{jk}| \text{ or } (x_{ik} - x_{jk}|)^2 \text{ for numeric-valued attributes} \\ &= 1 \text{ if } (x_{ik} \neq x_{jk}) \text{ and 0 otherwise for nominal variables} \end{aligned} \tag{6}$$

and s_k represents the average dissimilarity for attribute k,

$$s_k = n^{-2} \sum_{i}^{n} \sum_{j}^{n} \delta_{ijk},$$

Friedman and Muelman (2004) show how this framework can be extended to defining clusters that are determined from weights that may differ by both attribute and group so that (5) is replaced by a group-specific distance measure,

$$D_{ij}^{(k)} = \sum_{\ell=1}^{p} \omega_{\ell k} d_{ijk} \tag{7}$$

and the partition S_K and the selection of weights $\{\omega_{\ell k}\}$ are selected jointly to minimize (4), where in (3), D_{ij} is replaced by $D_{ij}^{(k)}$. Friedman and Muelman (2004) also provide algorithms and illustrations. This extension of the basic clustering approach is important because it provides a way of discovering cluster structure where some or all groups may only be discernable based on group-specific subsets of attributes.

Predictive Clustering

Clusterwise regression provides a way to find group-specific models. In its simplest form it does not provide a direct way of classifying customers, but does provide a description of the cluster characteristics along with within-cluster models for the dependent variable of interest. In its early form (Spath 1979, 1982), proposed that one jointly find the K-group partition of (3) and the corresponding set of within-cluster regression coefficients, so as to minimize the sum of within group sum-of-squared residuals of the K models:

$$C(S_K) = \sum_{k=1}^{K} \left(\underline{y}^{(k)} - \underline{Z}^{(k)}\underline{b}^{(k)}\right)' \left(\underline{y}^{(k)} - \underline{Z}^{(k)}\underline{b}^{(k)}\right), \tag{8}$$

where $\underline{y}^{(k)}$, $\underline{Z}^{(k)}$, and $\underline{b}^{(k)}$ represent the dependent variable, design matrix and regression coefficients for group k, k = 1, . . . K. DeSarbo, Oliver and Rangaswamy (1989), Wedel and Kistemaker (1989), and Wedel and Steenkamp (1989, 1991) are seminal papers that extended this approach and show how it could be applied to market segmentation. DeSarbo and Grisaffe (1998), DeSarbo and DeSarbo (2001), and Brusco, Cradit and Taschian (2003) have all proposed multicriterion forms of clusterwise regression. In the next section, we will also discuss how latent class models provide useful generalizations of the original approach.

CHAID (Chi-Squared Automatic Interaction Detection) provides another promising method of predictive clustering. Introduced by Kass (1980), algorithms have been developed by Magidson (1992, 1994) to provide an extremely powerful method of determining the predictive value of both nominal and ordinal predictors when the dependent variable is also of that type, and provides tree analyses that are generally more reliable than those provided by earlier binary tree procedures (including the first Automatic Interactive Detection procedures, although AID could also handle continuous variables). CHAID assesses the significance of all possible ways of using each predictor to subdivide the sample, and sequentially selects the most significant partition based on a chi-squared significance test that is adjusted to reflect that multiple tests are done on each predictor (i.e., the adjustment is based on the number of categories for each predictor). Magidson and Vermunt (2005) have also shown how this method can be extended to multiple dependent variables of mixed form (continuous, ordinal or nominal) by first fitting a latent class cluster model based on those multiple dependent variables, and then using the resulting latent classes as the univariate categorical variable in the CHAID analysis. We discuss latent class cluster analysis in the next section.

GENERAL LATENT CLASS MODELS FOR SEGMENTATION

This family of models provides a general way of developing segmentation frameworks that can include:

a. flexible classification functions with covariates and/or random effects;
b. general linear models with predictors, that may vary by segment, for representing complex within-segment relationships;
c. general features that accommodate within- and across-segment dependencies among the components of a multivariate dependent variable.

We start with a discussion of latent class cluster models, which provides a general way of fitting general linear models for a multivariate dependent variable within latent segments. We then look more specifically at within-segment models when we consider latent class regression.

Latent Class Cluster Models

In this model, the dependent variable is a vector of outcome measures, $\underline{y}_i = (y_{i1}, \ldots, y_{iJ})$ that are related directly to the marketing objectives. For example, these might include continuous measures of the incidence of purchase or the change in frequency of purchase of a brand of product or service, non-continuous indicators for the choice of brand or service, ordinal satisfaction measures, or even measures of this type for several related brands or services. It could include any combination of continuous, ordinal or nominal variables.

A very general approach, that subsumes many of the current latent class approaches, including latent class and "generalized mixture" regression (e.g., Wedel & Kamakura, 1998), was proposed by Vermunt and Magidson (2002). In this case, we find K segments by maximizing the J-variate likelihood for N observations

$$L = \prod_{i=1}^{N} \sum_{k=1}^{K} P(X_i = k | \underline{z}_i^{(cov)}, \Theta^{(1)})\, f_k(\underline{y}_i | \underline{z}_i^{(pred)}, \Theta_k^{(2)}), \qquad (9)$$

where $P(X_i = k | \underline{z}_i^{(cov)}, \Theta^{(1)})$ is the probability that a customer with covariate vector $\underline{z}_i^{(cov)}$ and parameters $\Theta^{(1)}$ is in latent segment (cluster) k, k = 1, . . ., K; and $f_k(\underline{y}_i | \underline{z}_i^{(pred)}, \Theta_k^{(2)})$ is the J-variate density within segment k, with predictors $\underline{z}_i^{(pred)}$ and parameters $\Theta_k^{(2)}$ (specific to segment k).

The Classification Probabilities

A standard approach is to model the logistic probability, $P(X_i = k | \underline{z}_i^{cov}, \Theta^{(1)})$, that observation *i* is in latent segment *k*, as the logistic probability function,

$$P(X_i = k | \underline{z}_i^{cov}, \boldsymbol{\Theta}^{(1)}) = \frac{\exp[\eta(k | \underline{z}_i^{(cov)}, \boldsymbol{\Theta}^{(1)})]}{\sum_{\ell=1}^{K} \exp[(\eta(\ell | \underline{z}_i^{(cov)}, \boldsymbol{\Theta}^{(1)})]}, \tag{10}$$

with

$$\eta(\ell | z_i^{(cov)}, \xi_j) = \gamma_{\ell 0} + \sum_{r=1}^{R} \gamma_{\ell r} z_{ir}^{(cov)}. \tag{11}$$

The Within Segment Model

Typically, when fitting these models, one initially assumes some form of local independence and, in the most extreme case, one could assume that the J components of y_i are completely independent conditional on a customer being in segment k,

$$f_k(\underline{y}_i | \underline{z}_i^{(pred)}, \boldsymbol{\Theta}_k^{(2)}) = \prod_{j=1}^{J} f_{kj}(y_{ij} | \underline{z}_i^{(pred)}, \boldsymbol{\Theta}_{kj}^{(2)}), \tag{12}$$

where the parameters $\Theta_{kj}^{(2)}$ for the univariate density f_{kj} will generally consist of a varying number of components depending on the mathematical form of the density f_{kj} (Vermunt & Magidson, 2002). Then, based on the analysis of bivariate residuals from an initial fit of (12), we can consider more realistic types of dependence, where say only L subgroups, $\underline{y}_i^{(\ell)}$, of the J components are independent, i.e., the subgroups $\underline{y}_i^{(\ell)}$, ℓ **= 1, . . ., L,** consist of multivariate, and possibly univariate, mutually exclusive subvectors, $\underline{y}_i^{(\ell)}$, of the original J outcome measures, such that:

$$(\underline{y}_i^{(1)}, \dots, \underline{y}_i^{(\ell)}, \dots, \underline{y}_i^{(L)}) = (y_{i1}, \dots, y_{iJ}). \tag{13}$$

In this case the K segments are found by maximizing the likelihood in (9), with

$$f_k(y_i \mid z_i^{(pred)}, \Theta_k^{(2)}) = \prod_{\ell=1}^{L} f_{k\ell}(\underline{y}_i^{(\ell)} \mid \underline{z}_i^{(pred)}, \Theta_{k\ell}^{(2)}) \tag{14}$$

The mutivariate normal distribution provides one of the easiest (and statistically manageable) ways of modeling dependence among variables. Consequently, we generally assume that the continuous components, $\underline{y}_i^{(\ell)}$, are normal or have been transformed so that they are approximately normal and that the corresponding density $f_{k\ell}$ in (14) is normal or multivariate normal. If the normal model is clearly inappropriate for a given component $y_i^{(k)}$ of the subvector of component measures $\underline{y}_i^{(\ell)}$, a simple approach is to transform using the inverse univariate cumulative normal distribution applied to the empirical distribution (or scaled ranks) of $y_i^{(k)}$, e.g.,

$$y_i^{(k)} = \Phi^{-1}\left(\text{Rank}(w_i^{(k)})/(N+1)\right) \tag{15}$$

where $w_i^{(k)}$ is the original continuous (but clearly non-normal) measure. This approach and the subsequent interpretation of the segmentation profiles is illustrated in Chaney, Cooil, and Jeter (2006).

In unusual cases, it is possible that the multivariate distribution of normal univariate components will not be multivariate normal. So, as an alternative approach, one could consider normalizing an entire subvector of component measures, $\underline{y}_i^{(\ell)}$, with an appropriate multivariate transformation. A special case of this latter approach would be to use normalized factors based on the original components.

When there are subvectors, $\underline{y}_i^{(\ell)}$, in (13) that consist of ordinal (or nominal) components, the density $f_{k\ell}$ in (14) would typically be a restricted joint multinomial distribution, where category probabilities are modeled as ordinal (or nominal) logistic regressions. In this way, each of the L subgroups, $\underline{y}_i^{(\ell)}$, of the original components of $\underline{y}_i$, consists of exclusively continuous or exclusively ordinal components. Dependence between continuous and ordinal (or nominal) components of $\underline{y}_i$ can be introduced by adding the ordinal variables to the group of predictors $\underline{z}_i^{(pred)}$ (Vermunt & Magidson, 2002). The need for such direct effects is also deduced by analyzing the residuals from a preliminary fit of model (12) and subsequent fits of model (14). Note however, that when ordinal (or nominal) variables, that are components of the vector of outcome measures, are also used as predictors for other outcome measures this will affect the way in which the within class models for each dependent component can be used. If these ordinal (or nominal) variables can be viewed only as outcome measures, whose values are to be predicted from within class models, they clearly cannot also be used as predictors. In this case it may be necessary to study groups of outcome measures separately.

Model Selection Procedure

A flexible approach is to start by including all plausible covariates for latent class (or segment) membership. One can then repeatedly apply the following two-step procedure until there is no further improvement according to an appropriate statistical criterion. Typically the Bayesian Information Criterion (BIC), or the Akaike Information Criterion are used (Akaike, 1974; Schwarz, 1976; Bozdogan, 1987). Step A focuses on the selection of the predictors and the covariances among dependent variables, while step B governs the selection of covariates for the classification function in (10).

> *Step A:* Given a specific selection of covariates, a forward stepwise procedure is used to select those covariances and predictors that would provide a better model (e.g., lower BIC); at each step, the best candidate covariances and predictors are determined by an analysis of Lagrange-type chi-square statistics based on the bivariate residuals of the outcome measures, $\underline{y}_i$ (Vermunt & Magidson, 2002; 2005), and the optimal number of segments is redetermined.
>
> *Step B:* Subject to the selection of predictors and covariances in step A, one could then use a backward-stepwise procedure to eliminate the classification covariates. At each step the optimal number of segments is redetermined.

By initially including all possible covariates, and first identifying the appropriate predictors and covariances (Step A), before eliminating those covariates that are not important, the procedure provides an auspicious starting point for finding the best model (e.g., minimum BIC) that is consistent with confirmatory residual diagnostics. DeSarbo, Kamakura, and Wedel (2005) discuss other model selection criteria that are appropriate in latent class analysis.

Adding Random Effects

One can easily generalize the classification function in (10) to accommodate random effects. This provides one way of modeling additional dependencies among observations. For example, imagine that we want to fit a longitudinal model for a vector of outcome measures taken from the same set of households over successive periods. Vermunt (2003) considers parametric and nonparametric models of this kind. Of course, continuous random effect models generally require much fewer parameters.

As an example of such a model, imagine we have data for J groups (e.g., households) and wish to consider K latent segments. Let $\underline{y}_{ij}$ represent the ith observed vector of measures in household j (i = 1,..., n_j). We find K segments by maximizing the likelihood for N observations, ($N = \sum_{j=1}^{J} \mathbf{n}_j$):

$$\log L = \sum_{j=1}^{J} \log \int_{\xi_j} \prod_{i=1}^{n_j} f\,(\underline{y}_{ij} \mid \underline{z}_{ij}^{(cov)}, \underline{z}_{ij}^{(pred)}, \xi_j, \boldsymbol{\Theta}) \quad f(\xi_j)\, d\xi_j\,), \tag{16}$$

with

$$f(\underline{y}_{ij} \mid \underline{z}_{ij}^{(cov)}, \underline{z}_{ij}^{(pred)}, \xi_j, \boldsymbol{\Theta}) = \sum_{k=1}^{K} P\,(X_{ij} = k \mid \underline{z}_{ij}^{(cov)}, \boldsymbol{\Theta}^{(1)})\ f_k\,(\underline{y}_{ij} \mid \underline{z}_{ij}^{(pred)}, \boldsymbol{\Theta}_k^{(2)}), \tag{17}$$

and $P(X_{ij} = k \big| Z_{ij}^{cov}, \xi_j, \Theta^{(1)})$ is the logistic probability that observation i, from household j, is in latent class k,

$$P\,(X_{ij} = k \mid \underline{z}_{ij}^{(cov)}, \xi_j, \boldsymbol{\Theta}^{(1)}) = \frac{\exp[\eta\,(k \mid \underline{z}_{ij}^{(cov)}, \xi_j\,)]}{\sum_{\ell=1}^{K} \exp[(\eta(\ell \mid \underline{z}_{ij}^{(cov)}, \xi_j)]}\,, \tag{18}$$

with

$$\eta(\ell \mid \underline{z}_{ij}^{(cov)}, \xi_j) = \theta_{\ell 0}^{(1)} + \sum_{r=1}^{R} \theta_{\ell r}^{(1)} z_{ijr}^{(cov)} + \theta_{\ell 00}^{(1)}\ \xi_j\,, \tag{19}$$

for $\ell = 1, \ldots, K$. In (18) and (19), the ξ_j, $j = 1, \ldots, J$, are assumed to be independent standard normal random effects, which is not very restrictive given the flexibility that is provided in (19) by the random effect coefficients, $\theta^{(1)}_{\ell 00}$, which correspond to each latent class ℓ, $\ell = 1, \ldots, K$, and which are subject to the identifiability constraint $\sum_{\ell=1}^{K} \theta^{(1)}_{\ell 00} = 0$.

Cooil, Keiningham, Aksoy and Hsu (in press) use this type of model to study changes in a household share of wallet measure (a univariate dependent variable). That paper identifies 14 customer segments and studies the within-segment effects of baseline satisfaction, changes in satisfaction, and other customer and business transaction variables. In that case a household random effect is used in the classification function (18).

Latent Class Regression Models (Mixture Regression)

This is just a special case of the latent cluster model described in (9) through (14), where the dependent variable is a single outcome measure. Nevertheless, by focusing on one outcome measure, it is possible to consider substantially more complicated within-segment models, and to use a larger number of coefficients that vary by segment. This frequently is not practical in cluster models because a large number of parameters are already needed to accommodate the multiple component dependent variable, and the dependencies among those components.

For a continuous dependent measure, typically one would use a general linear model, so that the within-segment probability function for segment k, $f_k(\underline{y}_i | \underline{z}_i^{(pred)}, \Theta_k^{(2)})$ of (9) or (17), is a univariate member of the exponential family, with link function $\eta(\mu_k)$ that is a linear function of the p predictors $Z_{ik} = (z_{ik1}, \ldots, z_{ikp})$ with coefficients $\Theta_k^{(2)} = (\beta_{k0}, \beta_{k1}, \ldots, \beta_{kp})$ for segment k,

$$\eta(\mu_k) = \beta_{k0} + \sum_{r=1}^{p} \beta_{kr} z_{ikr}^{(pred)}. \tag{20}$$

For example, if the dependent measure is continuous normal (or has been transformed to that it is approximately normal) the link function in (20) is just the identity, i.e., $\eta(\mu_k) = \mu_k$. If instead the dependent measure is ordinal (e.g., a satisfaction measure on a Likert scale), one could use an adjacent-category logit, so that within segment *k*, (20) becomes

$$\log\{P[Y_i = j+1 \mid \underline{z}_{ki}^{(pred)}, B_k] / P[Y_i = j \mid \underline{z}_{ki}^{(pred)}, B_k]\} = \beta_{kj0} + \sum_{p=1}^{P} \beta_{kp} z_{kpi}^{(pred)} \quad (21)$$

(here Y_i is the random variable associated with the dependent measure y_i for customer *i*.). The standard alternative formulation for (21) would be the proportional odds form of the cumulative logit, but typically the adjacent-category model tends to fit just as well (Agresti, 2002, p. 287), so that the choice between the two formulations is generally based on how one would prefer to interpret the coefficients. An advantage of (21) is that it provides an intercept and predictor coefficients that have a relatively straightforward managerial interpretation:

$\exp(\beta_{kj0})$ = ratio of the probability of being in category j + 1 relative to category j (Or the "risk of j + 1 relative to j"), *ceteris paribus,* when Y_i is in class k,

and

$\exp(\beta_{kp})$ = change in ratio of probabilities (j + 1 relative to j) per unit change of predictor z_{pk}, *ceteris paribus,* when Y_i is in class k.

A third, less practical, alternative to (21) would be to assume a logit function where predictor coefficients actually depend on the category j (e.g., in (5), substitute β_{jp} for β_p). Although this would provide a more flexible model, it obviously requires a considerably larger number of parameters, and the results would be much more difficult for a manager to use. Fortunately, in many empirical applications this more parametrically complex alternative does not provide competitive scientific models (in terms of BIC) relative to the adjacent-category formulation (21).

Selection Procedure for Latent Class Regression Models

For these models the general two step procedure for general cluster models is considerably easier to execute. In particular, Step A would focus only on the selection of predictors (while step B would be a search for covariates, conditional on the choice of predictors). Of course when there is a relatively modest set of candidate predictors and covariates, a global search for the best model may be feasible.

Empirical Examples

To illustrate latent class analysis, we apply these models to data collected as part of a market research study. Data for these examples came from a two-period study of customers who shop at a large international specialty-goods retail chain. In the first example, we used their overall satisfaction with the shopping experience at time 2 (measured on a five level Likert scale) as the univariate dependent variable $\underline{\mathbf{y}}$. The candidate covariates (and predictors) were: age, gender, overall satisfaction at time 1, and the baseline (time 1) satisfaction and the change in satisfaction (between times 1 and 2) on four service attributes:

- employee availability to assist them;
- check-out speed;
- cashier friendliness;
- employee knowledgeability and ability to answer questions.

For all satisfaction variables, the likert scale levels are: 1: "The Best," 2: "Better Than Most"; 3: "Average"; 4: "Worse Than Most"; 5: "The Worst." Age and gender were not significant as covariates or predictors.

Given the ordinal form of the dependent variable (overall satisfaction at time 2), latent class ordinal-logistic regression models are appropriate. The best scientific model (minimum BIC) of this type was fit to 2560 customer profiles. It consists of three customer segments, and uses the remaining nine customer variables (initial overall satisfaction, and the initial levels of satisfaction and changes in satisfaction for the four component services) only as covariates for determining how customers should be classified to segments. These variables are not effective predictors within segment. Due to missing values, the available number of customer profiles changes as the subset of customer variables changes. Thus, in contrast to definitions of BIC that are sometimes used in statistical packages, we were careful to define BIC as the average sample value,

$$\mathrm{BIC} \equiv [-2\log(\text{maximum likelihood}) + p\log(n)]/n, \qquad (22)$$

where n and p represent the available number of observations (customer profiles), and the number of parameters, respectively.

The best model has 26 parameter degrees of freedom. The within-segment models for $\underline{\mathbf{y}}$ are ordinal logistic regressions with only six parameter degrees of freedom. These parameters are for the intercepts of the five levels (four degrees of freedom) along with only two additional parameters for overall segment differences (two degrees of freedom). The classification component of the best model has 20 parameter degrees of freedom (18 degrees of freedom for coefficients of the nine covariates across three segments and two additional degrees of freedom for intercepts).

Among the covariates, the least significant are the variables *Employees Ready to Assist at Time 1*, and *Check-Out Speed at Time 1*, although both are still significant at the 0.05 level. All other covariates are significant at the 0.01 level. The classification error rate is 17%. Lambda (the proportional reduction in classification error relative to classifications made randomly to segments of the same size) and entropy R-squared are both 61%. When only the covariates are used (without knowledge of the dependent variable) the classification error rate is 30%.

The characteristics (profiles) of the three segments are summarized in Table 1. In terms of the dependent variable (overall satisfaction at time 2), we label these segments as "Satisfied" (69% rate at the "Better Than Most" level, and 27% at "The Best" level), "Very Satisfied" (99% at "The Best" level) and "About Average" (60% at the "Average" level, and 36% at "Better Than Most" level). The sizes of these segments are 56, 28, and 17% of the sample, respectively, and while all changes in overall satisfaction are positive, the change is largest in Segment 3 (the smallest, least satisfied group). Interestingly, in the two larger most satisfied groups, changes in component satisfaction are always negative, in contrast to Segment 3, where all changes in component satisfaction are positive except for *Cashier Friendliness.*

Using information from the same data set, we also studied latent class cluster analyses of a dependent variable $\underline{\mathbf{y}}$ with two continuous components:

- total customer expenditures after time 1 (up to and including time 2);
- change in overall satisfaction between times, expressed as a conditional percentile relative to all other customers with the same satisfaction level at time 1.

We considered the same selection of customer variables. In this case the best models generally require from 12 to 19 segments. In these models we also found that very few of the customer variables are effective within-segment predictors, and the same set of covariates were generally still important for classification.

CONJOINT ANALYSIS

In this final section on approaches to segmentation, we summarize methods that are appropriate in the context of conjoint analysis. This is a procedure that focuses on the estimation of buyer preferences for product attributes, where these preferences are inferred based on how respondents evaluate different product-attribute profiles. Possible attributes include not only functional and physical features, but also product characteristics such as brand name and price. In some cases these attribute levels may be presented as benefits, or described in symbolic or qualitative terms. Preference data are obtained in many different forms after various profiles of attribute levels are presented to subjects. A vast number of marketing models and statistical estimation techniques have been used in these analyses (Wedel & Kamakura, 1998, chap. 17).

Green and Krieger (1991) provide a conceptual framework (with empirical examples) for how various types of segmentation analyses fit into the context of a conjoint analysis. For example, a-priori segmentation studies may be conducted using segments that are defined in terms either a selected set of product-attribute profiles (the stimulus set) or by using individual attribute levels to define appropriate segments. In this case, logistic regression or other supervised classification methods are useful ways of studying how the selected profiles are related to other variables. Post-hoc approaches would use the preference data itself to determine the appropriate segments to target for product design (Green & Krieger, 1991, pp. 22-23) and in this case latent class models provide one of the most flexible frameworks for the analysis.

DISCUSSION

Market segmentation is clearly one of the most important concepts in marketing. In fact, the Migros case study illustrates how firms can construct novel and inventive approaches that provide great value. A-priori and custom designed post-hoc analyses are among the most important

TABLE 1. Description of Three Customer Segments

Segment Attributes	Segment		
	1	2	3
Segment Size (%)	56	28	17
Segment Size Based on Covariate Predictions	57	24	19
Reification	Satisfied	Very Satisfied	About Average
Dependent Variable: Overall Satisfaction at Time 2 (% at Each Level and Average)			
The Best (1)	27	99	0
Better Than Most	69	1	36
Average	4	0	60
Worse Than Most	0	0	2
The Worst (5)	0	0	1
Average Rating	1.8	1.0	2.7
Average Covariate Values			
Employees Ready to Assist (Time 1)	1.9	1.4	2.3
Change	-0.16	-0.16	0.45
Check-Out Speed (Time 1)	1.9	1.5	2.2
Change	-0.13	-0.11	0.37
Cashier Friendliness	1.7	1.3	2.1
Change	-0.12	-0.07	-0.28
Employees Are Knowledgeable & Answer Questions (Time 1)	1.8	1.4	2.2
Change	-0.21	-0.11	0.45
Overall Satisfaction (Time 1)	1.9	1.3	2.2
Average Values of Inactive Covariate			
Conditional Percentile of Change in Satisfaction [a]	72	49	92

Note. All change variables represent the change in the corresponding component of service satisfaction between times 1 and 2 (on the five point Likert Scale). The dependent variable and all other active covariates are measured on a five point Likert scale (1: Best, 5:Worst).

[a] The conditional percentile represents the percentile (or scaled rank) of change in satisfaction between periods among all customers who started at the same satisfaction level in period 1.

approaches that a firm should consider. Nevertheless, our review has focused on statistical methods that are flexible and useful in many different business and data environments. We also review some recently developed methods that provide very powerful tools for segmentation analysis.

For post-hoc descriptive analysis, new cluster algorithms are important exploratory tools for determining whether different combinations of customer variables should actually be used to define different segments (Friedman & Meulman, 2004). These tools may provide the only way of identifying certain types of customer segments. They also provide new ways of viewing the segmentation framework.

Latent class analysis is another important method for descriptive and predictive studies. It provides a direct way to simultaneously estimate the classification functions, and the within-class models for the dependent variables that define the objectives of the segmentation task. In fact, software is now available for a large family of models of this kind. Typically these program packages allow one to consider either univariate or multivariate dependent variables of mixed type. Multiple dependents make it possible to define segments in terms of multiple objectives. For example, these dependents may include information on customer profitability, or even measures of the quality of customer decision making (Aksoy, Bloom, Lurie, & Cooil, 2006). One can also fit multi-level models by including random effects as a component of the classification function. This additional flexibility is often needed when the observed customer responses are only conditionally independent within pre-defined groups, or strata (e.g., households, or time-periods).

Although the best statistical representation may provide substantial, statistically identifiable, stable and responsive segments, those same segments may not always be accessible, or actionable. This will happen whenever the model itself indicates that it is difficult to accurately classify individual customers to those segments. That is, statistically identifiable segments are not always completely identifiable in a marketing sense. In these cases, the analysis may provide a statistical definition of a large, clearly defined, and important segment, that is also responsive to specific marketing efforts, but to which it is difficult to accurately assign individual customers. This is frequently the case when group-random effects are needed in the best scientific model. This type of segmentation analysis is nevertheless useful and may, for example, still provide important information on the expected results of marketing campaigns for products that are not targeted to specific segments. This is especially true if the goal of the segmentation analysis is to determine the appropriate product line for a firm, and such a model may suffice if the firm makes all competitive items in its product line available to all buyers, as is often the case (Green & Krieger, 1991, p. 23). Finally, an analysis that statistically defines a subset of segments that are not completely identifiable or actionable (e.g., because of large random effects), may simultaneously identify another group of segments that do meet these criteria (Cooil et al., 2006).

The empirical examples of latent class analyses provide an illustration of how critical it is to select dependent variables that represent the objectives of the segmentation study. This must be done effectively and parsimoniously because the best model for a multivariate dependent variable may be extremely complex (as was the case with the second empirical example where the dependent variable had two components), and this additional complexity may be unnecessary if the dependent variables are poorly chosen. One way to avoid such problems is to start with exploratory latent class analyses that use inactive covariates. These may include alternative dependent measures or factors based on a preliminary set of dependent variables (e.g., several principal components). The measure labeled "Conditional Percentile of Change in Satisfaction," at the bottom of Table 1, is an example of an inactive covariate that provides an alternative way of studying the change in overall satisfaction over time. If it were included as an active covariate, the model itself would be logically circular and not very useful–final overall satisfaction (time 2) would be modeled in terms of overall satisfaction at time 1 and a measure of change in overall satisfaction–but as an inactive covariate, it provides a way of studying how the segments are related to another type of satisfaction measure. Chaney, Cooil, and Jeter (2006) also illustrate how principal components can be used as inactive covariates to provide an important alternative way of understanding segment characteristics.

AUTHOR NOTE

Bruce Cooil builds statistical models that are designed to help solve problems in health care and marketing. He has specific research interests in latent class and grade-of-membership models, qualitative data reliability, large sample estimation theory and extreme value theory. His publications have appeared in over 20 refereed journals in business, statistics and medicine, and have over 1200 citations.

Lerzan Aksoy's research interests include consumer satisfaction and its relationship to customers' spending patterns, customer relationship management, impact of recommendation agents in the online environment and consumer decision quality. She is co-author of the book *Loyalty Myths* with Keiningham, Vavra and Wallard published by Wiley. Her articles have been accepted by such publications as *Journal of Marketing*, *Journal of Service Research* and *Sloan Management Review*. Her articles have been nominated for best paper in *Journal of Service Research* and *Managing Service Quality*.

Timothy L. Keiningham is author of several management books and numerous scientific papers. His most recent book, *Loyalty Myths* (with Vavra, Aksoy, and Wallard), 2005 published by Wiley, poses the fallacies of most of the conventional wisdom surrounding customer loyalty. He has won best paper awards from the *Journal of Marketing* and the *Journal of Service Research*. His articles have appeared in such publications as *Journal of Marketing*, *Journal of Service Research*, *International Journal of Service Industry Management*, *Journal of Relationship Marketing*, *Managing Service Quality*, and *Journal of Retail Banking*.

REFERENCES

Agresti, A. (2002). *Categorical data analysis*. Hoboken, New Jersey: Wiley.

Akaike, H. (1974), A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19(6), 716-723.

Aksoy, L., Bloom, P.N., Lurie, N.H., and Cooil, B. (2006). Should recommendation agents think like people? *Journal of Service Research*, 8, 297-315.

Ansoff, H.I. (1957). Strategies for diversification. *Harvard Business Review*, Sept.-Oct.: 113-124.

Bijmolt, T.H.A., Paas, L.J., & Vermunt, J.K. (2004). Country and consumer segmentation: multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing*, 21, 323-340.

Bock, T. & Uncles, M. (2002). A taxonomy of differences between consumers for market segmentation. *International Journal of Research in Marketing,* 19, 215-224.

Bodapati, A.V. & Gupta, S. (2004). The recoverability of segmentation structure from store-level aggregate data. *Journal of Marketing Research*, 41, 351-364.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, 52, 345-370.

Brusco, M.J., Cradit, J.D., & Tashchian, A. (2003). Multicriterion clusterwise regression for joint segmentation settings: an application to customer value. *Journal of Marketing Research*, 40, 225-234.

Chaney, P., Cooil, B. & Jeter, D. (2006). A classification of firms based on earnings attributes. Unpublished manuscript, available at SSRN: http:/ssrn.com

Cheung, K., Kwok, J.T., Law, M.H., & Tsui, K. (2003). Mining customer product ratings for personalized marketing. *Decision Support Systems*, 35, 231-243.

Cheung, K., Tsui, K., & Liu, J. (2004). Extended latent class models for collaborative recommendation. *IEEE Transactions on Systems, Man and Cybernetics–Part A: Systems and Humans*, 34 (1).

Chung, J. & Rao, V.R. (2003). A general choice model for bundles with multiple-category products: application to market segmentation and optimal pricing for bundles. *Journal of Marketing Research*, 40, 115-130.

Cooil, Bruce, Timothy L. Keiningham, Lerzan Aksoy, and Michael Hsu (2007). A Longitudinal Analysis of Customer Satisfaction and Share of Wallet: Investigating the Moderating Effect of Customer Characteristics. *Journal of Marketing,* Vol. 71, no. 1 (January), 67-83.

DeSarbo, W., Kamakura, W., & Wedel, M. (2005). Latent structure regression. In R. Grover & M. Vriens (Eds.), *The Handbook of Marketing Research: Uses, Misuses, and Future Advances* (Chapter 19), Sage Publications.

DeSarbo, W.S. & DeSarbo, C.F. (2001). A generalized normative segmentation methodology employing conjoint analysis. In Anders Gustafsson, Andreas Herrmann, and Frank Huber, (Eds.), *Conjoint Measurement: Methods and Applications*, 2d. ed. Berlin: Springer, 447-478.

DeSarbo, W.S. & Grisaffe, D. (1998). Combinational optimization approaches to constrained market segmentation: an application to industrial market segmentation. *Marketing Letters*, 9 (2), 115-134.

DeSarbo, W.S. & Cron, W.L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification* 5, 249-282.

DeSarbo, W.S., Oliver, R.L., & Rangaswamy, A. (1989). A simulated annealing methodology for clusterwise linear regression. *Psychometrika*, 54, 707-736.

Du, Rex Y. & Kamakura, W.A. (2006). Household life cycles and lifestyles in the United States. *Journal of Marketing Research*, 43, 121-132.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179-188.

Foody, G.M., & Mathur, A. (2004). A relative evaluation of multiclass Image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1335-1343.

Friedman, J.H. (2006). Comment: Classifier technology and the illusion of progress. *Statistical Science*, 21, 15-18.

Friedman, J.H., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28, 337-407.

Friedman, J.H. & Meulman, J.J. (2004). Clustering objects on subsets of attributes. *Journal of the Royal Statistical Society,* Series B, 66, Part 4, 815-849.

Gower, J.C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27 (4), 857-871.

Green, P.E. (1977). A new approach to market segmentation. *Business Horizons*, 61-73.

Green, P. & Krieger, A. (1991). Segmenting markets with conjoint analysis. *Journal of Marketing*, 55, 20-31.

Gupta, S. & Chintaugunta, P.K. (1994). On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research*, 31, 128-136.

Hastie, T., Tibshirani, R. & Friedman, J.H. (2001). *The elements of statistical learning*. New York: Springer.

Hand, D. (2006). Classifier technology and the illusion of progress. *Statistical Science*, 27, 1-14.

Hogan, J.E., Lemon, K.N., & Rust, R.T. (2002). Customer equity management: charting new directions for the future of marketing. *Journal of Service Research*, 5 (1), 4-12.

Kamakura, W.A., Wedel, M., & Agrawal, J. (1994). Concomitant variable latent class models for conjoint analysis. *International Journal of Research in Marketing,* 11, 451-464.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29, 119-127.

Khan, R.J. & Jain, D.C. (2005). An empirical analysis of price discrimination mechanisms and retailer profitability. *Journal of Marketing Research*, 42, 516-524.

Kim, Y., Street, W.N., Russell, G.J., & Menczer, F. (2005). Customer targeting: a neural network approach guided by genetic algorithms. *Marketing Science*, 51 (2), 264-276.

Kotler, P. & Armstrong, G. (1999). *Principles of marketing*. Prentice Hall: London.

Krieger, A.M. & Green, P.E. (1996). Modifying cluster-based segments to enhance agreement with an exogenous response variable. *Journal of Marketing Research*, 33, 351-363.

Libai, B., Narayandas, D., & Humby, C. (2002). Toward an individual customer profitability model. *Journal of Service Marketing*, 5 (1), 69-76.

Magidson, J. (1992). Chi-squared analysis of a scalable dependent variable. *Proceeding of the 1992 Annual Meeting of the American Statistical Association, Section on Statistical Education*.

Magidson, J. (1994). The CHAID approach to segmentation modeling: chi-squared automatic interaction detection. In R.P. Bagozzi (Ed.), *Advanced Methods of Marketing Research*. (pp. 118-159). Cambridge, MA: Basil Blackwell.

Magidson, J., & Vermunt, J.K. (2005). An extension of the CHAID tree-based segmentation algorithm to multiple dependent variables. In C. Weihs and W. Gaul (Eds.), *Classification: The Ubiquitous Challenge*, 240-247. Heidelberg: Springer

McDonald, M. & Dunbar, I. (2004). *Market segmentation: how to do it, how to profit from it*. London: Elsevier.

Mittal, V. & Kamakura, W.A. (2001). Satisfaction, repurchase intent, and repurchase behavior: investigating the moderating effect of customer characteristics. *Journal of Marketing Research*, 38, 131-142.

Peterson, R.A. (1972). Psychographics and media exposure. 12 (June) *Journal of Advertising Research,* 12 (June), 17-20.

Ruiz, J. & Chabet, J., & Hansen, P. (2004). Another trip to the mall: a segmentation study of customers based on their activities. Les Cahiers du Gerard, G-2004-09.

Rygieksji, C., Wang, J., & Yan, D.C. (2002). Data mining techniques for customer relationship management. *Technology in Society*, 24, 483-502.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6: 461-464.

Sinha, I. & DeSarbo, W.S. (1988). An integrated approach toward the spatial modeling of perceived customer value. *Journal of Marketing Research*, 35, 236-249.

Späth, H. (1979). Algorithm 39: clusterwise linear regression. C*omputing,* 22, 367-73.

Späth, H. (1982). Algorithm 48: clusterwise linear regression. C*omputing,* 29, 175-81.

Srinivasan, V. & Park, C.S. (1997). Surprising robustness of the self-explicated approach to customer preference structure measurement. *Journal of Marketing Research*, 34, 286-291.

Verhoef, P.C., Spring, P.N., Hoekstra, J.C. & Leeflang, P.S.H. (2002). The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands. *Decision Support Systems,* 34, 471-481.

Vermunt, J.K. (2003). Multilevel latent class models. *Sociological Methodology*, 33: 213-239.

Vermunt, J.K. & Magidson, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced.* Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K. & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars and A. McCutcheon (Eds.). *Applied latent class models*, 89-106. Cambridge University Press.

Wedel, M. & Kamakura, W. (1998). *Market segmentation: conceptual and methodological foundations*. Boston: Kluwer.

Wedel, M. & Kistemaker, C. (1989). Consumer benefit segmentation using clusterwise linear regression. *International Journal of Research in Marketing,* 6, 45-59.

Wedel, M. & Kistemaker, C. (1989). A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research,* 28, (November) 385-96.

Wind, Y. (1978). Issues and advances in segmentation research. *Journal of Marketing Research*, 15, 317-337.

Wu, J. & DeSarbo, W.S. (2005). Market segmentation for customer satisfaction studies *via* a new latent structure multidimensional scaling model. *Applied Stochastic Models in Business and Industry*, 21, 303-309.

Xu, R. & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions of Neural Networks*, 16 (3), 645-678.

Zhong, P. & Fukushima, M. (2006). A new multi-class support vector algorithm. *Optimization Methods and Software*, 21, 359-372.

doi:10.1300/J366v06n03_02