Concomitant variable latent class models for conjoint analysis

Wagner A. Kamakura\textsuperscript{a}, Michel Wedel\textsuperscript{b,*}, Jagdish Agrawal\textsuperscript{c}

\textsuperscript{a} University of Pittsburgh, Pittsburgh, PA, USA
\textsuperscript{b} Department of Business Administration and Management Science, Faculty of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, Netherlands
\textsuperscript{c} California State University, Hayward, CA, USA

Final version received January 1994

Abstract

We propose a unifying framework for benefit and demographic segmentation based on the analysis of rank-order or choice data collected in conjoint studies. The model identifies a number of unobserved segments, estimates the conjoint model within each segment, and at the same time estimates the association of segment membership with concomitant (consumer descriptor) variables. Analyses of synthetic datasets, and an illustrative application of the model to a conjoint analysis study on consumers' preferences for banking services are provided. The analyses of synthetic data support the validity of the concomitant variable latent class model. The validity of the model is also confirmed in an empirical study, where the preference patterns are shown to be logically consistent with the respondents' banking behavior.

1. Introduction

Considerable research interest has been devoted recently to theoretical issues concerning market segmentation in conjoint analysis (Green and Krieger, 1991). In commercial applications, market segmentation ranks among the primary purposes for performing conjoint analysis, both in the US (Wittink and Cattin, 1989) and in Europe (Wittink et al., 1993). The focus of these conjoint studies has been mostly on benefit segmentation, where segments are formed on the basis of common preferences, so that products or services can be optimally designed and targeted. Another important issue, however, is the accessibility of benefit segments; once segments are identified and products designed to suit their tastes, the manager has to be able to identify members of the segments so that marketing efforts can be directed to them. In other words, a benefit segmentation scheme has to be related to observables characteristics of its members, in order to be actionable.

The model presented in this paper addresses several of the important issues concerning market segmentation. This model allows the manager to simultaneously identify benefit segments, estimate a probabilistic rank or choice model within each segment, and to relate each benefit segment to observable characteristics of its members.

\textsuperscript{*} Corresponding author.
1.1. Market segmentation with conjoint analysis

Both in academia and in practice, segmentation in conjoint studies has traditionally been performed using a priori segmentation schemes (Green, 1977), or using the so-called two-stage procedure (Green and Srinivasan, 1978, 1990). In the a priori scheme, the researcher first defines groups of consumers a priori, based on their observable characteristics (e.g., socio-demographics) and then estimates the aggregate preference function of all consumers within each group. The two-stage procedure starts by first estimating the preference functions for each individual consumer in the sample. These estimated functions are then used as the criteria to form clusters of consumers with similar preferences or tastes. The limitations of these traditional procedures have been documented by e.g., Moore (1980), Steenkamp and Wedel (1993), DeSarbo et al. (1992), and Elrod et al. (1992). The success of a priori segmentation depends heavily on the strength of the relationship between the observable characteristics used to form the a priori groups, and consumer preferences or choices. But the relationship between e.g. socio-demographic variables (typically used in a priori segmentation) and preferences is generally too weak to explain differences across consumers. Two-stage segmentation procedures, on the other hand, ignore the estimation error associated with the part-worth estimates from the first stage, and are outperformed by the clusterwise regression class of methods (Kamakura, 1988; Wedel and Kistemaker, 1989; Wedel and Steenkamp, 1989).

The development of conjoint segmentation techniques that alleviate the limitations of the traditional procedures has received much interest recently. Methods have been proposed by Hagerty (1985), Kamakura (1988), Ogawa (1987), Wedel and Kistemaker (1989), Wedel and Steenkamp (1989, 1991), DeSarbo et al. (1989), DeSarbo and Cron (1988), DeSarbo et al. (1992), and DeSarbo et al. (1992). Reviews of the relative advantages and disadvantages of these methods have been provided by Wedel and Steenkamp (1989), Green and Helsen (1989), DeSarbo et al. (1992) and are not reiterated here. We describe two of the above methods, i.e. those proposed by Ogawa (1987) and DeSarbo et al. (1992), in some detail, as they are pertinent to and provide a motivation for the development of the model we propose.

DeSarbo et al. (1992) proposed a latent class specification for simultaneously performing conjoint analysis and segmentation, tailored to metric data. The model involves mixtures of multivariate conditional normal distributions, and employs a stochastic framework (as opposed to the deterministic nature of most previous conjoint segmentation models), which allows for (asymptotic) significance tests of the part-worths within the segments. The method also allows for posterior probabilities to be derived that indicate consumers' membership in the identified segments. In an empirical application of the method DeSarbo et al. attempted to relate these posterior probabilities to a number of demographic variables but no significant associations were found in this second step of the analysis. The authors demonstrate empirically that their procedure outperforms the traditional two-stage procedure.

One of the most common uses of conjoint analysis is the prediction of choices or choice shares for a new product. However, due to their deterministic nature, the commonly used conjoint models are not well suited for choice share predictions. Researchers interested in predicting the market share for a new product based on the conjoint model must thus resort to arbitrary choice simulators that attempt to translate regression estimates into choice probabilities. Thus far, the only conjoint segmentation method that is tailored to the analysis of rank order preferences in a probabilistic framework (thus producing estimates of choice probabilities) is the one proposed by Ogawa (1987). His formulation employs a stochastic logit framework. He proposes a ridge-like procedure for estimating individual level part-worths, and an information theoretic criterion to aggregate consumers (hierarchically). The method, however, employs individual-level estimates at the first steps of the aggregation. As documented by Wedel and Kistemaker (1989) and Elrod et al. (1992), this two-stage approach can result in potentially unreliable identification of benefit segments, due to the instability and
bias in the individual-level part-worth-estimates. Further, to circumvent the computational burden involved in estimating a logit model at each stage of the clustering algorithm, Ogawa uses OLS to estimate part-worths, which decreases the efficiency of these estimates.

Building upon the work of Ogawa (1987) and DeSarbo et al. (1992), we develop a conjoint segmentation model which is tailored to the analysis of rank order preferences or choice-data. Rank-order preferences are the second most frequently collected data in commercial conjoint studies (Wittink and Cattin, 1989). Recently, there has also been much scientific interest in conjoint analysis of choice data, because it can be directly used in predicting actual choice behavior and choice shares. In a recent study, Elrod et al. (1992) indicated the need to “incorporate heterogeneity explicitly in choice models in the not too distant future, which would eliminate a major concern about estimating choice models entirely at the aggregate level” (p. 376).

The development of such models is the issue of the present work. The models are developed using a stochastic framework, and involve mixtures of multinomial distributions (Kamakura and Russell, 1989). Consumers are grouped into segments and part-worths are simultaneously estimated within each segment assuming a multinomial logit model. The model thus integrates the DeSarbo et al. (1992) segmentation procedure for metric data with procedures for the analysis of conjoint studies in which rank-ordered preferences or choice data are collected, using aggregate multinomial logit models (Louviere, 1988). Moreover, our model generalizes the DeSarbo et al. (1992) procedure by simultaneously relating consumers’ segment membership to specified socio-economic and demographic (concomitant) variables. It thus avoids an analysis of the posterior probabilities of segment membership in a second and unrelated step, thereby increasing the power to predict segment membership. Our model is based upon recently developed latent class models, proposed by Dayton and Macready (1988) and Formann (1992). In these models, the probabilities of latent class membership are functionally related to concomitant variables. Gupta and Chintagunta (1992) applied such models to the analysis of scanner data.

Our model and its estimation procedure will be detailed in the next section. Section 3 contains an analysis of two synthetic datasets which supports the performance of the model. Section 4 contains an application to a study on the usage of services by customers of a bank. The predictive performance of the model with concomitant variables is shown to exceed that of the model without concomitant variables. The last section contains conclusions.

2. A latent class probabilistic conjoint model with concomitant variables

The purpose of the analysis is to estimate a model that: 1) relates consumers’ rank ordering or choices of a set of profiles to predetermined attributes of these profiles, 2) identifies a number of latent classes in which the part-worths of this conjoint model differ, and 3) at the same time estimates the association between consumers’ segment membership and a number of consumer descriptor (concomitant) variables. The model assumes that consumers rank or make choices that offer the maximum unobservable utility among the available alternatives. This utility is assumed to have a measurable component (as a function of the characteristic of the choice alternative) and a random utility component (McFadden, 1974). Let:

\[ i = 1, \ldots, n \] consumers,
\[ j = 1, \ldots, J \] profiles,
\[ k = 1, \ldots, K \] attributes,
\[ l = 1, \ldots, L \] concomitant variables describing each consumer
\[ s = 1, \ldots, S \] latent classes,
\[ X_{jk} \] the \( k \)th independent conjoint variable for profile \( j \),
\[ Z_{il} \] the \( l \)th concomitant (descriptor) variable for consumer \( i \).

For rank order preference data \(^1\), we assume

\(^1\) For choice data, the observed choice variable \( y_{it} \) is defined as 1 if consumer chooses alternative \( j \) at choice occasion \( t(=1,2,\ldots,T) \) and 0 otherwise.
that the ranking process is directional, and that a rank order may be considered as a set of successive and independent first choices, in which the respondent first chooses the alternative with the highest unobservable utility among all available, then chooses the most preferred among the remaining ones, and so on. Now let:

\[ t = 1, \ldots, T \]  

\[ y_{ijt} = 1 \] if consumer \( i \)'s rank order of stimulus \( j \) is greater or equal to \( t \), and 0 otherwise.

We assume that the observed choice variables \( y_{ijt} \) arise from a population composed by \( S \) unobserved segments, where the a priori probability that a consumer comes from segment \( s \) is \( \theta_{s|Z} \), which is conditional on the characteristics \( Z \) of the consumer, as explained later. Given segment \( s \), the probability of the choice of stimulus \( j \) at choice replication \( t \), is:

\[ P_{jt|s} = \text{Prob}\left[U_{jt|s} \geq U_{qt|s}\right], \tag{1} \]

where \( U_{jt|s} \) is the random utility derived from alternative \( j \) at \( t \) by members of segment \( s \), and \( U_{qt|s} \) is the maximum utility among the alternatives ranked at position \( t \) or lower \(^2\).

\[ U_{qt|s} = \max\{U_{jt|s}, U_{qt-1|s}\} \tag{2} \]

The formulation in (1) and (2) implies that at each rank consumers choose the profile that has maximum utility over the remaining set. This formulation is equivalent to the rank-explosion rule derived by Chapman and Staelin (1982) from Luce and Suppes (1965), as well as to Ogawa's (1987) rank logit model. The random utility is assumed to be a function of the known attributes of the choice alternative \( j \):

\[ U_{jt|s} = \beta_{0js} + \sum_{k=1}^{K} \beta_{kjs} x_{jk} + \epsilon_{jts}. \tag{3} \]

The random components, \( \epsilon_{jts} \), are assumed to be i.i.d. Weibull, leading to a multinomial logit model with rank probabilities \(^3\) (Chapman and Staelin, 1982):

\[ P_{jt|s} = \frac{\exp[U_{jt|s}]}{\sum_{q \in Q} \exp[U_{qt|s}]}, \tag{4} \]

where \( Q \) is the set of alternatives ranked lower or equal to \( j \). It is now assumed that the prior probability that consumer \( i \) comes from latent class \( s \) is a function of the concomitant (descriptor) variables, \( Z_{it} \). We formulate the model for the prior probabilities, which is called the submodel (Dayton and Macready, 1988) as:

\[ \theta_{s|Z} = \frac{\exp\left(\sum_{l=1}^{L} \gamma_{ls} Z_{il}\right)}{\sum_{s=1}^{S} \exp\left(\sum_{l=1}^{L} \gamma_{ls} Z_{il}\right)}, \tag{5} \]

where \( \gamma_{ls} \) is a parameter that denotes the impact of the \( l \)th consumer characteristic on the prior probability for class \( s \), and \( \Sigma \gamma_{ls} = 0 \). The submodel described above formulates the prior probability of membership to the segments as a logistic function of consumer characteristics. The parameters of this multinomial logit submodel are specific to each descriptor variable and market segment; a positive \( \gamma_{ls} \) implies that a higher value of descriptor \( Z_{it} \) increases the prior probability that a consumer belongs to segment \( s \). Alternative forms of the submodel in concomitant variable latent class models are discussed by Dayton and Mcready (1988).

Eqs. (3)–(4) provide the conditional (on membership to a particular segment) choice or rank probabilities. The unconditional probability that consumer \( i \) chooses profile \( j \), at (rank order or choice replication) \( t \) is obtained by combining these conditional probabilities with the prior membership probabilities from Eq. (5):

\[ P_{jt} = \frac{1}{S} \sum_{s=1}^{S} \theta_{s|Z} P_{jt|s}, \tag{6} \]

\(^2\)For choice data, it is assumed that the consumer maximizes utility over the entire choice set (McFadden, 1974).

\(^3\)Choice probabilities would be given by (McFadden, 1974):

\[ P_{jt|s} = \frac{\exp[U_{jt|s}]}{\sum_{l=1}^{L} \exp[U_{lt|s}]}. \]
Eq. (6) shows that the unconditional choice probabilities can be decomposed into a weighted average of latent choice probabilities \( P_{ij|s} \), where the weights \( \theta_{ij|Z} \) vary systematically as a function of the concomitant variables. Eq. (4)–(6) thus provide an indirect link between the consumers' characteristics and their choice probabilities. This relationship can be used to assess the effect of individual characteristics upon preferences and choice probabilities, as demonstrated in our empirical illustration.

Estimates of the parameters of the model, \( \beta_{k} \), and the parameters of the submodel, \( \gamma_{ls} \), are obtained via maximum likelihood, as described in the Appendix. Once estimates of the parameters are obtained, the posterior probability that consumer \( i \) belongs to latent class \( s \) can be calculated, using Bayes' rule:

\[
\alpha_{i|s} = \frac{\theta_{ij|Z} L_{i|s}}{\sum_{s=1}^{S} \theta_{ij|Z} L_{i|s}},
\]

where:

\[
L_{i|s} = \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ij|s}^{X_{it}},
\]

denotes the likelihood of consumer \( i \), conditional upon being in class \( s \). The posterior segment-membership probabilities \( \alpha_{i|s} \) can be interpreted as a fuzzy allocation of consumer \( i \) to the benefit segments \( s = 1, \ldots, S \). Previous research (De Sarbo et al., 1992) have attempted to relate these posterior allocations to the demographic characteristics of each consumer, in a two-stage approach. In contrast, our approach makes an explicit linkage between the segments and consumers' demographic background in a single stage, via the concomitant submodel (Eq. 6).

2.1. Determining the number of segments

One of the major difficulties in market segmentation is the determination of the number of segments. Inferences about the number of segments is one of the issues that has received the least satisfactory statistical treatment in the literature. Formal tests for the number of segments, such as the likelihood ratio test, cannot be applied because the asymptotic properties of these tests do not hold (c.f., Aitkin and Rubin, 1985; Titterington, 1990). A number of procedures have been proposed to determine the number of segments such as the use of Akaike's Information Criterion – AIC (Akaike, 1974) and the Consistent Akaike Information Criterion (CAIC) (Bozdogan, 1987), and the use of Monte Carlo significance tests (Aitkin et al., 1981; De Soete and DeSarbo, 1991). These latter tests, however, involve extensive computational effort, and are not feasible for the large datasets often encountered in marketing research. The CAIC and AIC statistics, on the other hand, are burdened with the same problems as the likelihood ratio statistic in their use for determining the number of segments, being grounded on asymptotic expansions assuming the regularity conditions also used in the derivation of the likelihood ratio test. Therefore they can be used only as heuristics to guide the determination of the possible number of underlying segments, where the number of segments selected has the minimum value of these statistics.

More recently, researchers have focused their attention to new criteria that use the estimated information matrix (Windham and Cutler, 1992; Bozdogan, 1993). In our study, we use one of these new criteria (ICOMP), developed by Bozdogan (1993). This entropic statistical complexity criterion extends Akaike's Information Criterion by adding a correction for model complexity that is measured by the complexity of the estimated inverse information matrix. By including this correction, ICOMP controls for the risks of either being overly parsimonious or of over-parameterizing a model. This new criterion can be easily computed as,

\[
\text{ICOMP} = -2\max(t) + c\ln(\text{tr}(\Sigma)) - \ln(|\Sigma|),
\]

where

\( t = \) total number of parameters in the model,
\( \Sigma = \) covariance of the estimates, calculated as the inverse of the information matrix.

The first component of the ICOMP criterion,
(-2max), measures the lack of fit of the model, while the last two terms measure the complexity of the covariance of the estimates (calculated as the inverse of the information matrix). Therefore, ICQ measures the balance between improved fit with a more saturated model (more segments) and the increased complexity of such a model (see Bozdogan, 1992, for a detailed discussion).

3. Analysis of synthetic data

In this section we apply the model to two synthetic data sets to illustrate its performance in recovering “true” model and submodel parameters in a sample. Our purpose is to show that the segment-level preference functions and the components of the concomitant submodel can be all estimated simultaneously, given data on the preference rankings (or choices) and the descriptor variables from a sample of consumers. In order to verify that the estimation procedure is indeed capable of uncovering the “true” parameters of the model, we first generate synthetic data using a known set of parameters, and then demonstrate that one cannot reject the hypothesis (with $\alpha = 0.05$) that the estimates obtained from the model are equal to the “true” parameters used to generate the data.

The data generated in this small simulation study are multinomial choices among profiles by subjects in a number of segments, the memberships being defined by concomitant variables. In order to create the concomitant variable and choice data, we first define three segments $s=1, 2, 3$, with the utility functions defined in (3). We assume four profiles, $j = 1, 2, 3, 4$, and two attributes $(k = 1, 2)$. The prior probability of membership of a consumer in a segment is defined by the function:

$$
\ln \frac{\theta_{s|z}}{\theta_{s|\bar{z}}} = \gamma_{oz} + \sum_{l=1,2} \gamma_{lz} Z_{lz} + \delta_{lz},
$$

where $\gamma_{oz}$ is a segment-intercept, $\gamma_{lz}$ denotes the impact of the $l$th consumer characteristic on the prior probability, and $\delta_{lz}$ are i.i.d. extreme-value errors that are assumed to account for omitted

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates from two synthetic data sets</td>
</tr>
<tr>
<td>Attribute</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\beta_{1s}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\beta_{2s}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Concomitant variables

| $\gamma_{0s}$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| $\gamma_{1s}$ | -0.164 | -0.441 | -0.007 | -0.057 |
|          | (0.213) | (0.481) | (0.152) | (0.171) |
| $\gamma_{2s}$ | 0.2 | -0.2 | 0.0 | 0.5 | -0.5 |
|          | 0.152 | -0.065 | 0.535 | -0.329 |
|          | (0.107) | (0.182) | (0.117) | (0.133) |
| $\gamma_{3s}$ | -0.5 | 0.5 | 0.0 | -1.0 | 1.0 |
|          | -0.456 | 0.680 | -0.959 | 0.952 |
|          | (0.132) | (0.257) | (0.139) | (0.155) |

1. "True" value used to generate the choice data.
2. Parameter estimate.
variables. The synthetic data were generated by first specifying "true" values for the coefficients $\beta_k$ and $\gamma_l$ for each segment (see Table 1 below). A sample of 500 hypothetical consumers was then created with values of the attributes $X$ and concomitant variables $Z$. Each hypothetical consumer was classified in one of the three segments by drawing a random value from a Weibull distribution, and then computing the membership probabilities, $\theta_{ij}$ defined in (10). Each consumer was assigned to the segment with the highest probability of membership.

The same process was used (except that utilities were computed using Eq. (3)) to generate 10 choices for each consumer, based on the coefficients for the segment to which the consumer was assigned. This procedure was used to generate choice and concomitant variable data for two sets of known "true" values for the model and sub-model parameters, under two different scenarios.

In the first scenario, segment allocations and discrete choices were generated under "noisy" conditions, i.e., the stochastic components were large relative to the deterministic components of the utility functions and the latent class submodels. The second scenario reflects a more deterministic choice behavior, and a more accurate definition of prior membership probabilities based on demographics. Table 1 compares the "true" values with the estimates obtained with the MLE procedure described earlier for multinomial choice data, for each of the two scenarios.

The results in Table 1 clearly show the high degree of randomness in the choice data under the 'high noise' condition as the $R^2$ equals 4.7%, ($R^2$ is computed as one minus the ratio of the maximum log-likelihood for the current model to the likelihood under equal choice probabilities and prior membership probabilities). The log-likelihood of the model was $-6128.8$. Most importantly, the parameter estimates (especially the $\beta_k$) are reasonably close to the actual values used to generate the data, even under these severe conditions. One cannot reject ($p < 0.05$) the hypothesis that the estimated parameters equal the true values. Moreover, the results displayed in Table 2 show that the estimated model is able to correctly classify between 59% to 86% of the members of the three latent classes, even under this high degree of randomness in the choice data.

The $R^2$ of the second scenario was 21.9%, and the log-likelihood of the model under this scenario was $-5024.0$, reflecting the low noise condition (lower than scenario 1). Once again, one cannot reject ($p < 0.05$) the hypothesis that the estimated parameters equal the true values. As expected, the model's ability to correctly classify the members of each latent class based on the observed choices is higher under this second scenario, as shown in Table 2.

This analysis of synthetic data illustrates the ability of our model to recover true parameters and to classify consumers correctly into segments. The purpose of this limited analysis of synthetic data was to provide an illustration of the performance of the method. For further evidence of the performance of concomitant variable latent class models in general we refer to the original work of Dayton and Macready (1988) and Formann.

| Table 2 |
| Segment classification in the synthetic data (500 cases) |

<table>
<thead>
<tr>
<th>Scenario 1: Segment membership</th>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>136 (59%)</td>
<td>13 (6%)</td>
<td>83 (36%)</td>
<td>232 (100%)</td>
</tr>
<tr>
<td>Group B</td>
<td>7 (2%)</td>
<td>292 (76%)</td>
<td>83 (22%)</td>
<td>382 (100%)</td>
</tr>
<tr>
<td>Group C</td>
<td>39 (10%)</td>
<td>14 (4%)</td>
<td>333 (86%)</td>
<td>386 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario 2: Segment membership</th>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>220 (95%)</td>
<td>2 (1%)</td>
<td>10 (4%)</td>
<td>232 (100%)</td>
</tr>
<tr>
<td>Group B</td>
<td>1 (0%)</td>
<td>359 (94%)</td>
<td>22 (6%)</td>
<td>382 (100%)</td>
</tr>
<tr>
<td>Group C</td>
<td>38 (10%)</td>
<td>6 (1%)</td>
<td>342 (89%)</td>
<td>386 (100%)</td>
</tr>
</tbody>
</table>
Table 3
Statistics for the 1 to 5 segment solutions of the models with and without concomitant variables

<table>
<thead>
<tr>
<th>Segment</th>
<th>Without concomitant variables</th>
<th>With concomitant variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log-L</td>
<td>$R^2$</td>
</tr>
<tr>
<td>1</td>
<td>-2479.2</td>
<td>21.9%</td>
</tr>
<tr>
<td>2</td>
<td>-2185.5</td>
<td>31.2%</td>
</tr>
<tr>
<td>3</td>
<td>-2070.4</td>
<td>34.8%</td>
</tr>
<tr>
<td>4</td>
<td>-2009.2</td>
<td>36.7%</td>
</tr>
<tr>
<td>5</td>
<td>-1984.6</td>
<td>37.5%</td>
</tr>
</tbody>
</table>

(1992), who have examined the performance of other forms of concomitant latent class models.

4. Application

4.1. Description of the conjoint study

In this section our model will be applied to a commercial conjoint study on consumers’ preferences for bank services. Recently, a large bank in the U.S. was interested in knowing how valuable were different service attributes to its customers, and in identifying benefit segments among its customers, so that a menu of checking accounts could be created to best serve them. The bank’s managers wanted information on consumers’ concern with four attributes, each of which was varied at three levels in the study. These four attributes (with their levels in parentheses) were:

- **MINBAL:** the minimum balance required to exempt the customer from a monthly service fee ($0, $500, $1000);
- **CHECK:** the amount charged per check issued by the customer (0c, 15c, 35c);
- **FEE:** the monthly service ($0, $3, $6);
- **ATM:** the availability and cost of automatic teller machines in a network of supermarkets (not available, free ATM, 75c per transaction).

Two equivalent but distinct sets of profiles were created from the attributes, using a fractional factorial design (Addelman, 1962), one of which was used in the estimation of our model, while the other served as a hold-out sample to assess predictive validity. The two sets of nine profiles were presented to a random sample of 269 of the bank’s customers, in the form of “peeling stickers” in a mail survey. The customers were instructed to peel off their first choice and stick it to a designated place, then their second choice and so on, until a full ranking was obtained. (Note that this task enforces successive first choices to be made, which conforms to the rank choice theorem). The profiles were presented in a randomized order. In addition to the conjoint data, the following information for each respondent was obtained from the bank:

- **BALANCE:** average balance kept in the account during the past 6 months, earning 5.5% interest;
- **NCHECK:** number of checks issued per month in the past 6 months (at no charge);
- **NATM:** number of ATM transactions per month (all ATM machines).

These variables represent the actual past banking behavior of each respondent, and constitute the concomitant variables included in the model to explain segment membership.

4.2. Empirical results

As the dependent variable collected in the conjoint study consists of rank ordered preferences, we used the rank order choice model for the analyses. The effects of the first three at-

---

4 While a complete sample of 314 customers was obtained in the survey, 269 had been customers of the bank for a full year. Since we wanted to relate their preferences to their past banking experience, this subsample of 269 customers was used in our analysis.
Table 4
Parameter estimates of the 4-segment solutions of the latent class models

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Model without concomitant variables</th>
<th>Model with concomitant variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Seg. 1</td>
<td>Seg. 2</td>
</tr>
<tr>
<td>Attributes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MINBAL</td>
<td>-0.380</td>
<td>-1.978</td>
</tr>
<tr>
<td>FEE</td>
<td>-0.332</td>
<td>-0.171</td>
</tr>
<tr>
<td>ATM</td>
<td>0.524</td>
<td>2.421</td>
</tr>
<tr>
<td>ATM – 75c</td>
<td>-0.058</td>
<td>0.746</td>
</tr>
<tr>
<td>comitant variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BALANCE</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NCHECK</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NATM</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Size (%)</td>
<td>21.8</td>
<td>20.8</td>
</tr>
</tbody>
</table>

1 Statistically significant at \( p = 0.05 \)

Attributes, MINBAL, CHECK, and FEE, were assumed to be linear and were modelled using a single variable, while non-linear effects of the availability and cost of automatic teller machines were modelled by using two effect-type dummy-variables: ATM, indicating the free automatic

**ESTIMATED PARTWORTHS**

![Fig. 1. Estimates of the part-worths by segment.](image-url)
teller machines, and ATM-75c, indicating automatic teller machines available at 75c per transaction. The rank orders of the nine stimuli in the estimation set provided by the 269 respondents were analyzed with the concomitant variable latent class conjoint model detailed above, as well as a latent class conjoint model that did not include the concomitant variables, for comparison.

These models were applied to the data for \( S = 1 \) to 5 segments. Table 3 depicts for each of the solutions the results of the log-likelihood, the \( R^2 \) measure for goodness of fit, and the ICOMP criterion. The ICOMP-criterion reaches a minimum at \( S = 4 \) segments for both models. For the model without the concomitant variables \( R^2 \) was 36.7%, for the model with concomitant variables \( R^2 \) was somewhat higher, 37.5%. The ICOMP-criterion for the model without concomitant variables is somewhat lower than that for the model with concomitant variables, but since the former model is nested under the latter (by setting the coefficients for the concomitant variables to zero), a likelihood-ratio test can be used to compare these two models. A chi-square value of \( (2009.2 - 1983.6) = 51.2 \) with 12 degrees-of-freedom shows that the contribution of the concomitant variables to fit is statistically significant. Table 4 presents the estimated coefficients for the 4-segment solution of both models. As the estimated coefficients \( \beta_k \) of the models are quite similar, we only describe the results for the model with concomitant variables (Note that segment "size" in the concomitant variable model refers to the expected size, i.e. average posterior membership probabilities). Fig. 1 presents the partworths implied by the estimates reported in Table 4.

Fig. 1 shows that segment 1 displays significant effects of MINBAL, CHECK, FEE, and ATM. As compared to the other segments, customers in this segment are mainly concerned with the amount charged per check (CHECK) and the monthly service fee charged (FEE). Relatively little attention is paid to the minimum balance required (MINBAL) and the availability of automatic teller machines (ATM). Table 4 shows that the parameters for the submodel are all statistically significant. The positive coefficients for \( \text{BALANCE} \) and \( \text{NCHECK} \) indicate that customers with high average balance (BALANCE) and who issue a large number of checks (NCHECK) are more likely to belong to this segment than to others. This result for the submodel is consistent with the low concern for minimum balance of this segment and its high sensitivity to the amount charged per check. The negative coefficient for \( \text{NATM} \) in the submodel indicates that customers who are heavy users of ATM's are less likely to belong to this segment, which is consistent with their low concern for the availability of these machines.

Compared to the other segments (see Fig. 1) customers in segment 2 are the most sensitive to the availability of automatic teller machines in supermarkets (ATM, ATM-75c). Members of this segment assign a high value to free access to ATM's, and would rather pay 75c per ATM transaction than not have access to them. This is consistent with the effects of the concomitant (descriptor) variables in the submodel (Table 4), from which it may be observed that \( \text{NATM} \) (number of ATM transactions in the past year) was the only descriptor variable with a statistically significant coefficient for this segment. This positive coefficient for \( \text{NATM} \) indicates that heavy users of ATM's are most likely to belong to segment 2.

Customers in segment 3 are somewhat sensitive to minimum balance (MINBAL), and ATM-availability. The positive and statistically significant coefficient for \( \text{BALANCE} \) for this segment in the concomitant submodel (see Table 4) indicates that customers with high average balance are more likely to belong to this segment than to segments 2 or 4.

Table 4 and Fig. 1 show that segment 4 has a coefficient for the minimum balance required (MINBAL) nearly four times as large as that in segment 3. Although the other attributes also display significant coefficients, consumers in this segment are predominantly concerned with the minimum balance required to prevent a monthly service fee. The negative coefficient for \( \text{BALANCE} \) in the submodel explain this result: customers with low average balance have the highest probability of belonging to this segment.

The results of the proposed concomitant vari-
able model have high face validity. All coefficients for the preference functions displayed in Fig. 1 can be very well interpreted. The four segments all prefer lower balance requirement, low costs per check and the availability of (free) automatic teller machines in supermarkets, as expected. Most customers attach positive values only to free ATM’s, but the segment with the highest preference for these machines also ranks alternatives with 75¢ per transaction higher.

Moreover, the results of the concomitant submodel for each latent class are quite consistent with the preferences uncovered by the conjoint model in that segment. Consumers who use a large number of checks per month are concerned with the costs per check, consumers with a higher use of ATM value that service more than other consumers, and consumers with a lower average balance express a higher concern for minimum balance requirements. While these conclusions might seem rather obvious, they serve to attest the validity of the concomitant submodel.

Most importantly, the results provided by the model can be helpful to the bank managers, in the development of strategies best suited to its customer base. First, a number of segments were identified that differ in the importance placed on various service attributes. These results suggested a differentiated strategy with respect to the services offered. Moreover, our concomitant-variable conjoint model can be helpful in targeting. Based on the customers’ past banking behavior, the submodel provides predictions of the customers’ membership to each preference segment. This prediction, combined with the part-worth estimates within each segment, allows the manager to design the banking service most suitable to each particular customer. For example, a heavy user of checks will most likely be a member of segment 1, and will prefer banking services that best satisfy the preference function displayed in Fig. 1 for that segment. This customer will want a low cost per check, but will not mind keeping a large ($1000) minimum balance in the account to “pay” for these services. On the other hand, a customer who maintains a small monthly balance in the account, will be a likely member of segment 4, and prefer an account with low minimum balance, even if (s)he had to compensate for that by paying for ATM transactions or for every check issued (see Fig. 1). The viability of these differentiated services, would obviously depend on the marginal revenue generated by each customer, and the marginal costs of the service features.

4.3. Validity tests

In addition to the face validity discussed in the previous section, we also investigated the predictive validity of the concomitant variable latent class model relative to 1) a naive model, in which no attributes are considered, 2) an aggregate rank-order-logit model (e.g., the $S = 1$ solution), and 3) a latent class rank order logit model without concomitant variables. In assessing the predictive validity we used the preference ranking of the validation set of nine profiles, and calculated the values of the log-likelihood and $R^2$ (defined as in Section 3) on the basis of the parameters obtained from the estimation set. The results are depicted in Table 5. As one would expect, the goodness of fit (in the estimation sample) increases as more parameters are estimated, and is in general better than the predictive fit (in the validation sample). Table 5 also shows that the predictive fit obtained with the concomitant variable model is somewhat better than the one obtained with the simpler models, thus indicating its predictive validity. One must note, however, that the improvement in predictive fit with the concomitant submodel is only marginally better than the one obtained from the latent class rank-logit model without concomitant variables, despite the

<table>
<thead>
<tr>
<th>Sample Model</th>
<th>Parameter</th>
<th>Estimation sample</th>
<th>Validation sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>log-$l$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>NV</td>
<td>0</td>
<td>-3174.7</td>
<td>-</td>
</tr>
<tr>
<td>RL</td>
<td>5</td>
<td>-2479.2</td>
<td>0.219</td>
</tr>
<tr>
<td>LC</td>
<td>24</td>
<td>-2009.2</td>
<td>0.367</td>
</tr>
<tr>
<td>CV</td>
<td>36</td>
<td>-1983.6</td>
<td>0.375</td>
</tr>
</tbody>
</table>
statistical significance of the coefficients for the concomitant variables reported on Table 4.

5. Conclusions

The latent class probabilistic conjoint model proposed in the present paper has two important features. First, it provides a stochastic framework for segmentation of both rank-ordered and choice data collected in conjoint experiments, thus allowing for probabilistic predictions of choice events, or direct predictions of choice shares at the segment level. This simultaneous segmentation and estimation of conjoint models within segments under a stochastic framework also produces significance tests of the part-worths within segments. The model thereby builds upon the advantages of aggregate choice models noted by Elrod et al. (1992), while alleviating the limitation of such models caused by the assumption of preference homogeneity across consumers. By simultaneously identifying groups of consumers with similar preferences and estimating their preference functions, our model allows for preference heterogeneity across segments. A current limitation of the model is related to the independence of irrelevant alternatives (IIA) assumption, and the absence of interactions between attributes. However, IIA is assumed only within each benefit segment. The latent-class nature of the model in fact allows for independence among alternatives across segments (Kamakura and Russell, 1989).

Second, this model simultaneously relates consumers' segment membership to concomitant descriptor variables. This feature of the model alleviates the problem in previous latent class procedures of analyzing the posterior probabilities in a second and unrelated step, which often yields insignificant results (DeSarbo et al., 1992). This second feature is of critical importance for managerial decisions, as it makes it possible to identify customer profiles in each segment and to relate these profiles to the customers' preferences or partworths. Results from such a model, when combined with cost and revenue information, could be used to design services tailored to each specific segment, and to determine the target market (e.g., customer profile) for each of these services. From a substantive point of view, the proposed model explicitly deals with the distinction between the "active" variables in conjoint segmentation studies (an instance of the powerful approach of benefit segmentation (e.g., Wind, 1978)), and the "passive" or descriptor variables which are used to profile the segments in order to make them accessible. Our empirical analysis has demonstrated moreover that the inclusion of these concomitant (descriptor) variables in the conjoint model actually enhances the predictive power of the model, albeit by a marginal amount.

6. Appendix: Maximum likelihood estimation of the model

The likelihood function for the model is:

\[ L = \prod_{i=1}^{n} \sum_{s=1}^{S} \theta_{s|Z} \prod_{t=1}^{T} \prod_{j=1}^{J} P_{jt|s}^{y_{jt}} \]  

(A1)

Estimates of the parameters are obtained by setting the partial derivatives of the log likelihood to zero. The gradients of the log-likelihood required for the estimation are obtained as follows. The log-likelihood is:

\[ l = \sum_{i=1}^{n} \ln \sum_{s=1}^{S} \theta_{s|Z} L_{i|s} \]  

(A2)

Setting the derivative with respect to \( \beta_{ks} \) to zero yields:

\[ \frac{\partial l}{\partial \beta_{ks}} = \sum_{i=1}^{n} \frac{\theta_{s|Z}}{\sum_{s=1}^{S} \theta_{s|Z} L_{i|s}} \frac{\partial L_{i|s}}{\partial \beta_{ks}} \]

\[ = \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{j=1}^{J} \alpha_{it} (y_{ijt} - P_{jt|s}) X_{jk} \]  

(A3)

Eq. (A3) can easily be seen to be the ordinary stationary equation of the logit model for choice-, or rank order-data, fitted across all observations, where each observation on consumer \( i \) contributes with weight \( \alpha_{it} \).
Setting the derivative with respect to $\gamma_{ls}$ equal to zero yields:

$$
\frac{\partial l}{\partial \gamma_{ls}} = \sum_{i=1}^{n} \frac{\partial \ln \left( \sum_{s=1}^{S} \theta_{s|Z} L_{is} \right)}{\partial \theta_{s|Z}} \frac{\partial \theta_{s|Z}}{\partial \gamma_{ls}} = 0 .
$$

(A4)

Taking the derivative in (A4) and summing over $s$ gives:

$$
\sum_{s=1}^{S} \sum_{i=1}^{n} (\alpha_{ls} - \theta_{s|Z}) Z_{il} = 0 ,
$$

(A5)

It can be seen that Eq. (A4) is the ordinary stationary equation of a multinomial logit model, with the unobserved posterior probabilities $\alpha_{ls}$ as the dependent variables, and the consumer characteristics $Z_{il}$ as the independent variables. Obtaining the ML estimates thus amounts to solving the two sets of non-linear Eqs. (A3) and (A5).

The likelihood of latent class models can be maximized basically in two ways: by using the EM algorithm (Dempster et al., 1977), or by standard optimization routines such as the Newton-Raphson method. It is as yet unclear which of the two methods is to be preferred in general (see e.g., McLachlan and Basford, 1988; Everitt, 1984), but for models with nonlinear constraints such as the one at hand the Newton-Raphson procedure appears to have computational advantages (Moolaj and van der Heijden, 1992). In our solution of the maximum-likelihood problem we therefore use a modified Newton search with a trust-region global strategy (Dennis and Schnabel, 1983), based on the gradients given in the Appendix and the Hessian for the log-likelihood function.

References


