A Latent Class Binomial Logit Methodology for the Analysis of Paired Comparison Choice Data: An Application Reinvestigating the Determinants of Perceived Risk

Michel Wedel
Department of Business Administration and Management Science, Faculty of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

Wayne S. DeSarbo
School of Business, University of Michigan, Ann Arbor, MI 48109

ABSTRACT

A latent class model for identifying classes of subjects in paired comparison choice experiments is developed. The model simultaneously estimates a probabilistic classification of subjects and the logit models' coefficients relating characteristics of objects to choices for each respective group among two alternatives in paired comparison experiments. A modest Monte Carlo analysis of algorithm performance is presented. The proposed model is illustrated with empirical data from a consumer psychology experiment that examines the determinants of perceived consumer risk. The predictive validity of the method is assessed and compared to that of several other procedures. The sensitivity of the method to (randomly) eliminate comparisons, which is important in view of reducing respondent fatigue in the task, is investigated.

Subject Areas: Market Segmentation, Risk & Uncertainty, and Statistical Techniques.

INTRODUCTION

We introduce a model for identifying classes of subjects that exhibit differences in choice behavior in paired comparisons experiments. The method of paired comparisons was originally introduced by Fechner [14] and extended by Thurstone [23]. This data collection method is most often utilized in situations in which (N) objects are to be compared or judged qualitatively, or where quantitative judgements of preference cannot be easily made. Application areas of research where such a procedure is applied are discussed by David [8] and Dillon, Kumar, and Smith de Borrero [12].

Paired comparison tasks are popular in practice because of their simplicity. The data in paired comparison experiments are obtained by presenting all possible pairs of objects to one or more subjects, and each are required to choose the most preferred object of each pair. Incomplete designs to reduce the number of comparisons are also available [3]. The data thus obtained consist of binary responses with $N(N-1)/2$ distinct number of pairs or observations for each subject. We will focus on the situation where data on a number of descriptor variables are also available

Decision Sciences
Volume 24 Number 6
Printed in the U.S.A.

for each of the objects to be compared. In the statistical analyses of such data, the purpose is to relate these descriptor variables to the pairwise preference judgements. Carroll [5] has coined the term "external analysis" for this process.

External models are used for relating pairwise preference judgements to a specified configuration of the objects. In contrast with internal models, external analyses typically involve fewer parameters to estimate in describing differences in preferences. Such a procedure is the PREFPAIRS regression based procedure [2], which employs least-squares estimation and deals with individual differences by performing the analysis by subject. Often, in a subsequent step of the analysis, consumers are clustered on the basis of their estimated coefficients and the estimation is performed across subjects and profiles within segments. Wedel and Kistemaker [25] have shown, however, that the individual level analyses and the clustering procedures have lower (predictive) validity than procedures that simultaneously perform segmentation and estimation.

Another approach has been proposed by Cooper and Nakanishi [7] who employ logit models for the external analysis of paired comparisons data and deal with individual differences through the aggregation of the data into a priori defined homogeneous groups. This is an important limitation, as defining such homogeneous groups on the basis of subject demographic data may be problematic in many applications where such background variables are weakly related to such choice judgements.

Recently, procedures have been proposed that alleviate the problems of a priori definition of groups, individual level estimation, and two-stage procedures by simultaneously grouping subjects into a number of classes and estimating logit models of choice within each class. These models do not require a priori information on the group structure of subjects, but use a latent class formulation to derive a probabilistic group structure post-hoc. Such procedures have been proposed by Kamakura and Russell [18], who developed a latent-class multinomial logit model; by De Soete and DeSarbo [11], who developed a latent class probit model for the analysis of "pick-any out of N" data; and by Kamakura [17], who proposed a multinomial model that accommodates multimodal distributions of ideal points. (Related latent class models for internal analysis have been proposed [10] [12].)

THE LATENT CLASS BINOMIAL LOGIT MODEL

Model Formulation

Let:

i = 1, ..., I subjects;
j,k = 1, ..., N objects;
l = 1, ..., L descriptor variables of the objects;
X_{jl} = the l-th descriptor variable for the j-th object;
y_{ijk} = 1 if object j is preferred to k by subject i;
0 otherwise;
s = 1, ..., S latent classes or groups.

Suppose that I subjects are presented with $N(N-1)/2$ paired comparisons of distinct pairs of objects described by a set of common descriptor variables or design matrix ($\mathbf{X}=((X_{jl}))$), with response y_{ijk} denoting if subject i prefers object j to object k.

Assume that there exist S classes, and that y_{ijk} is binomially distributed. Each subject i belongs to one and only class s, which is not known in advance. The unconditional probability that subject i belongs to class s is denoted by α_s, where,

$$0 < \alpha_s < 1, \sum_{s=1}^{S} \alpha_S = 1. \tag{1}$$

Subject i is thus presented with two objects, j and k, and is asked to choose one of the pair preferred (assume no ties for now). Then, conditional upon subject i belonging in class s, the probability that subject i selects object j over k is:

$$P_{i|s}(y_{ijk} = 1) = P_{i|s}(V_{js} > V_{ks}), \tag{2}$$

where V_{js} is the latent utility of object j for any subject i in class s. Note that this utility is assumed identical for all subjects in class s. We assume that the sth class utility function can be represented as:

$$V_{js} = \sum_{l=1}^{L} \beta_{ls} X_{jl} + \varepsilon_{js}, \tag{3}$$

where

β_{ls} = the impact coefficient of the lth attribute on the derived utility in class s, and

ε_{js} = a random error component, which is independently and identically distributed with a weibull (or extreme value) distribution, that is:

$$P(\varepsilon_{js} < \varepsilon) = \exp(-\exp(\varepsilon)). \tag{4}$$

Then it follows that

$$P_{i|s}(y_{ijk} = 1) = P[\varepsilon_{ks} - \varepsilon_{js} < \sum_{l=1}^{L} \beta_{ls} X_{il} - \sum_{l=1}^{L} \beta_{ls} X_{kl}], \tag{5}$$

and thus

$$P_{i|s}(y_{ijk} = 1) = \frac{\exp [\sum_{l=1}^{L} \beta_{ls}(X_{jl} - X_{kl})]}{1 + \exp [\sum_{l=1}^{L} \beta_{ls}(X_{jl} - X_{kl})]} = P_{jk|i\varepsilon s}(\cdot). \tag{6}$$

Now, having obtained an expression for the probability that subject i prefers/chooses object j to k conditional upon being in class s, the conditional likelihood function for subject i (assuming independence over all distinct j,k pairs) can be formulated as:

$$L_{i|s} = \prod^{N}_{j<k}\prod P_{jk|i\varepsilon s}(\cdot)^{y_{ijk}}(1 - P_{jk|i\varepsilon s}(\cdot))^{1 - y_{ijk}}. \tag{7}$$

The unconditional likelihood for subject i is then:

$$L_i = \sum_{s=1}^{S}\alpha_s\prod^{N}_{j<k}\prod P_{jk|i\varepsilon s}(\cdot)^{y_{ijk}}(1 - P_{jk|i\varepsilon s}(\cdot))^{1-y_{ijk}}), \tag{8}$$

and the complete likelihood over the entire sample of subjects is:

$$L = \prod_{i=1}^{I}\sum_{s=1}^{S}\alpha_s\prod^{N}_{j<k}\prod p_{jk|i\varepsilon s}(\cdot)^{y_{ijk}}(1 - P_{jk|i\varepsilon s}(\cdot))^{1 - Y_{ijk}}, \tag{9}$$

or the log-likelihood:

$$\ln L = \sum_{i=1}^{I}\ln\left[\sum_{s=1}^{S}\alpha_s\prod^{N}_{j<k}\prod P_{jk|i\varepsilon s}(\cdot)^{y_{ijk}}(1 - P_{jk|i\varepsilon s}(\cdot))^{1-y_{ijk}}\right]. \tag{10}$$

By maximizing $\ln L$ with respect to α_s and β_{ls}, estimates of class membership and the associated impact coefficients for the attributes within classes can be obtained simultaneously. Once estimates of α_s and β_{ls} are obtained, each subject i can be assigned to each class s through the estimated posterior probability (using Bayes' rule) via:

$$\hat{\theta}_{is} = \hat{\alpha}_s\hat{L}_{i|s} / \sum_{s=1}^{S}\hat{\alpha}_s\hat{L}_{i|s}. \tag{11}$$

The maximum likelihood estimates of α_s and β_{ls} can be obtained by maximizing the likelihood function in (10) subject to the constraints in (1). This optimization problem can be solved by means of an EM-algorithm [9] [19]. Details of this algorithm are provided elsewhere [24].

When applying the above model to real data, the actual number of classes is typically unknown. Our approach to determine the appropriate number of classes involves the use of the consistent akaike information criterion (CAIC) [4] as a test heuristic:

$$CAIC = -2\ln L + n(S)(\ln(I \cdot N(N-1)/2) + 1), \tag{12}$$

where $n(S)$ is the effective number of parameters estimated in an S-class solution:

$$n(S) = L \cdot S + S - 1. \tag{13}$$

The CAIC can only be used as a heuristic, where one selects the value of S which yields a minimum CAIC, because the conditions for the theoretical justification for the use of CAIC (or a likelihood ratio) as a test statistic are not strictly met [21]. To ensure that the centroids of the classes are sufficiently separated from one another, we also use an entropy-based measure to examine the separation of the posterior probabilities of class-membership:

$$E_S = 1 + \sum_{i=1}^{I} \sum_{s=1}^{S} \hat{\theta}_{is} \ln \hat{\theta}_{is} / I \ln(S). \tag{14}$$

E_S is bounded between 0 and 1, where a value close to 1 implies that the classes are well separated.

A MONTE CARLO STUDY OF ALGORITHM PERFORMANCE

To assess the performance of our model and its estimation algorithm, we conducted a small Monte Carlo study. Seven independent factors were specified as having potential effects on algorithm performance. All factors were varied at two levels. The factors were chosen to represent conditions that are hypothesized to affect the performance of the method, the levels were chosen to provide meaningful differences between the alternative conditions. The factors (and their levels) were:

1. Number of subjects (32, 60);
2. Number of stimuli (8, 12);
3. Number of attributes (3, 7);
4. Number of classes (2, 4);
5. Error (mis)specification (weibull, normal);
6. Error variance (low, high);
7. Class separation (low, high).

The synthetic data were generated as follows. For a given number of classes (factor 4), the attribute coefficients for the first class were specified in the interval [-1.5, 1.5]. To specify classes that differ in the values of the coefficients, in the low class separation conditions (factor 7) coefficients in class 2 were obtained by adding .2 to the coefficients of class 1, the coefficients in class 3 were obtained by adding .2 to the coefficients in class 2, etc. In the high class separation condition, differences in the coefficients in successive classes were two times as high and obtained by adding .4 to the coefficients of the previous class. Subjects were assigned to classes in equal numbers.

The attributes were generated from a uniform distribution with the error component from a weibull or normal distribution (factor 5). This factor was included to investigate the effect of error (mis)specification: the weibull distribution represents the correct specification of the error leading to a binomial logit, while the normal distribution represents misspecified error, leading to a binomial probit. In the high error condition, the variance of the error component (factor 6) was multiplied by 4 relative to the low error condition. Increasing the relative level of error by a factor 4 is hypothesized to negatively affect parameter recovery. For each class, the utility function was calculated for each stimulus according to (3), and choices for each of the $N(N-1)/2$ pairs of stimuli were generated from (2).

Five dependent measures of algorithm performance were specified measuring parameter recovery, computational effort, and goodness of fit. These measures were:

1. Number of major iterations required for convergence: ITER;
2. log-L with actual parameters minus final log-L: L_DIF;
3. Root mean square error between β and $\hat{\beta}$: RMS(β);
4. Root mean square error between λ and $\hat{\lambda}$: RMS(λ);
5. Percentage of correct predictions: %PRED.

The seven independent factors were varied according to a 1/8 replicate 2^7 fractional factorial design [6] in 16 trials. The latent class methodology was applied to each of the 16 datasets. The five dependent measures of performance were calculated and analyzed with analysis of variance to examine the main effects of the independent factors. The adjusted mean values of the dependent variables for each of the factor levels are presented in Table 1. (Due to their skewed distributions ITER, RMS(β), and RMS(λ) were transformed by logarithms before the analyses, the logit transformation was applied to %PRED).

Computational requirements, as measured by the number of major iterations, is affected significantly only by the number of classes estimated: a larger number of classes increases the computational burden (Table 1). The difference in log-likelihood, as a measure of the goodness of fit, is affected only by the error (mis)specification, indicating a somewhat consistent fitting of the model over a variety of conditions. Increasing the number of subjects or stimuli results in an improvement in the estimation of the parameters β and λ (the effects of the number of subjects on RMS(λ) being nonsignificant).

This finding is consistent with general statistical theory on the effects of sample size on parameter estimation. Further, increasing the number of classes significantly decreases parameter recovery for both types of parameters. Again, this is consistent with general theory that increasing the number of parameters to be estimated decreases their recovery. Recovery of the parameters β is affected significantly by the error variance: increasing the error variance decreases the root mean squared error of β. Prediction accuracy was affected by the number of attributes; a larger number of attributes increases the percentage of correct predictions. This is probably caused by a larger variance of the systematic part of the utility function relative to the variance of the random component for increasing numbers of attributes, and does not mean that in practical applications predictions improve as the number of attributes increases. In applications increasing the number of attributes results an increasing complexity of the task for consumers may also affect their judgements. Prediction accuracy was also affected by the error variance (higher variance decreases prediction). Overall, parameter recovery is satisfactory and the estimates appear to be relatively robust to error misspecification. The percentage of correct predictions is over 80 percent across a variety of conditions.

APPLICATION: THE MEASUREMENT OF THE COMPONENTS OF PERCEIVED RISK

Background

The concept of perceived risk has been a focus of attention in the consumer psychology literature since Bauer [1] characterized consumer choice in terms of

Table 1: Means of the Monte Carlo study by factor levels.

Factor	Level	ITER$	LOG-L	RMS(β)$	RMS(λ)$	%PRED
Subjects	32	16.777	-8.1	.085[a]	.076	.845
	60	18.357	10.0	.054	.039	.860
Stimuli	8	15.959	6.9	.085[a]	.089[b]	.847
	12	19.298	-5.0	.054	.027	.858
Attributes	3	19.886	8.5	.061	.050	.808[c]
	7	15.487	-6.6	.075	.064	.896
Classes	2	8.499[c]	-2.8	.038[c]	.025[b]	.839
	4	36.234	4.7	.122	.091	.866
Error	weibull	17.993	25.2[c]	.079	.073	.855
	normal	16.945	-30.2	.056	.038	.849
Variance	low	16.610	9.3	.051[b]	.048	.892[c]
	high	18.728	-7.4	.091	.066	.812
Separation	low	21.328	1.1	.063	.064	.842
	high	14.440	.8	.073	.051	.863

$ geometric means after log-transformation
[a] $p<.10$
[b] $p<.05$
[c] $p<.01$

risk taking/reducing behavior. Subsequent research has tended to focus on two aspects of risk: the consequences/importance of a loss and the likelihood of unfavorable outcomes. (See [13] [21] [22] for comprehensive reviews of the perceived risk literature.) Jacoby and Kaplan [16] have identified six types of perceived risk associated with a product: performance, financial, safety, social, psychological, and time/opportunity. Peter and Ryan [20] later added opportunity/time loss to the preceding five components of risk.

Havlena and DeSarbo [15] report of a study examining the impact of these six components on perceived risk for drivers/owners of sports cars. We choose to utilize this study in order to augment the Havlena and DeSarbo [15] results in illustrating the proposed latent class binomial logit methodology. This methodology adds to the insights obtained from the approach used by [15] by providing the capability of simultaneously deriving groups or market segments of consumers who have different weights or importances for these various components of perceived risk—an aspect ignored in the authors' article.

In the study, 50 subjects were asked to make paired comparison judgements on 120 pairs of sports cars. Sixteen profiles were constructed using a fractional factorial, main-effects, conjoint design [1] using 13 attributes. These attributes were chosen to reflect the six categories of risk discussed previously. The attributes and their levels are depicted in Table 2. For further details of the study, refer to [15].

Table 2: Havlena and DeSarbo [15] design factors and levels.

Factor	Definition	Levels 1	Levels 2	Hypothesized Effect on Overall Risk for Level 2	Hypothesized Type of Risk Associated
Price	The base sticker price of the sports car	$20K	$40K	+	Financial
Warranty	The length of the warranty in years/miles	3yr/30K	6yr/60K	-	Financial
Country of Manufacture	Whether the manufacturer was located in Japan or the USA	USA	Japan	-	Performance
Advertising Endorsement	Whether a celebrity endorses the sports car or it is endorsed by *Road & Track* magazine	Celebrity	*Road & Track*	-	Performance
Safety Test Results	The analysis of the National Highway Traffic Safety Administration crash-test injury data as published in *Consumer Reports*	Not approved	Approved	-	Safety
Newness of Brand	Whether the brand of sports car, not the manufacturer, is new for the model year or whether the brand has been in existence for at least one year	New	Existing	-	Performance, psychological, or social
Manufacturer Reputation	Whether the manufacturer is best known for high performance or styling in their sports cars	Performance	Styling	+/-	Performance, psychological, or social
Repair Service	The length of time required for dealer service for most common repairs & maintenance	Same day	Next day	+	Opportunity
Loaner Car Availability	Whether or not the dealer makes a loaner car available for long term repairs	No	Yes	-	Opportunity
M.P.G.	The average mileage rating based on published test results	20	40	-	Financial
Special Financing	Whether the dealer offers reduced interest rates for financing the purchase of a sports car	None	1.9% rate	-	Financial

Table 2: (continued).

Factor	Definition	Levels 1	Levels 2	Hypothesized Effect on Overall Risk for Level 2	Hypothesized Type of Risk Associated
Braking System	Whether the sports car has standard disk brakes or computer-controlled anti-lock brakes to prevent skidding	Standard	Anti-lock	-	Safety
Position in Product Line	Whether the sports car is the standard model or the luxury/top-of-the-line model	Standard	Luxury	-	Performance, psychological, or social

Results

For comparison, the data for the elicited pairwise judgments of overall risk (i.e., overall, which of the two sports cars is riskier to buy) were analyzed with a traditional binomial logit approach. Table 3 presents the parameter estimates and standard errors of this aggregate logit model fitted across all subjects and stimulus-pairs. Note that these results are congruent with the Havlena and DeSarbo [15] aggregate PREFPAIRS regression analysis, presented in their Table 2.

The issue is whether all subjects exhibit the same perceived risk coefficient profile, or whether distinct groups of subjects exhibit different profiles. We applied the latent-class binomial logit model for $S=1$ to six classes. According to the minimum CAIC-rule, the $S=4$ class solution most parsimoniously represents the structure in the data. Table 4 presents the estimates of the parameters and asymptotic standard errors of the four class solution.

The first latent class, with a mixing proportion of .200, displays the largest coefficient for advertising endorsement, country of manufacture, newness of the brand, repair service, braking system, manufacturer reputation, and position in product line among the four latent classes. However, price, warranty, safety test results, and miles per gallon are also significant. This latent class most resembles the aggregate results among the four classes. Performance factors tend to obtain the largest of the coefficients in this latent class as compared to the remaining three.

In contrast, relative to the other three classes, latent class 2, with $\alpha_2=.160$, possesses the largest coefficients for warranty and special financing, although price, safety test result, manufacturer's reputation, and miles per gallon are also significant. The predominant concern for this group appears to be financially related, although other aspects of perceived risk are also significant. Interestingly, this is the only class that perceives cars made in America as less risky.

The third latent class, with $\alpha_3=.319$, possesses the highest coefficient on price (which is sizable), and miles per gallon as compared to the other three classes, although coefficients for warranty, country of manufacture, safety test results, newness of the brand, manufacturer reputation, and braking system are also significant. Here, the price coefficient is nearly four times the size of its closest rival (in latent class 2), and nearly three times the size of the aggregate solution.

Table 3: Binary logit model results ($S=1$).

Variable	Level	Aggregate
Price	= \$40K	1.303[c]
Warranty	= 6yr/60K	-.453[c]
Country of manufacture	=Japan	-.512[c]
Advertising endorsement	=Road & Track	-.099[a]
Safety test results	=Approved	-.998[c]
Newness of brand	=Existing	-.557[c]
Manufacturing reputation	=Styling	.275[c]
Repair service	=Next day	.175[c]
Loaner car availability	=Yes	-.046
Miles per gallon	=40	-.376[c]
Special financing	=1.9% rate	-.159[c]
Braking system	=Anti-lock	-.175[c]
Position in product line	=Luxury	-.144
Intercept		.113
Mixing proportion		1.000

[a] $p<.05$
[b] $p<.01$
[c] $p<.001$

Finally, the fourth latent class, with $\alpha_4=.321$, displays the highest coefficient for safety test results, as compared to the other classes. The coefficient for warranty, country of manufacture, newness of the brand, and miles per gallon are also very significant. Perceived risk in this class is predominantly based on safety.

The posterior probabilities for each of the 50 subjects are close to zero or one, which indicates good separation of the conditional mixture centroids. The value of the entropy measure for the four-class solution is $E_s=.994$, which indicates near perfect separation of the classes.

The results of the study clearly indicate the multidimensional nature of perceived consumer risk, and is thus congruent with the previous literature [15]. The three dominant dimensions of perceived risk that emerge for this product class are financial, safety, and performance risk. The analysis provides evidence that there is considerable heterogeneity in the sensitivity of subjects to these three major risk factors. The significant effects were consistently in the hypothesized directions in each of the classes.

The analysis also revealed that heterogeneity in the weights of the variables among segments of subjects may mask significance of the effects at the aggregate level. Position in product line was not significant at the aggregate level, but luxury/top of line model reduced risk relative to the standard model in classes 1 and 2, constituting 36 percent of the sample. An important implication of our findings is that a limited number of controlled risk reducing-strategies or perceived risk profiles can be devised, optimally matching the structure of risk perception in each of the classes.

Thus, the results of our latent-class binary logit model are consistent with the findings of the stochastic multidimensional scaling procedure applied by Havlena

Table 4: Parameter estimates of the four-class solution of the latent class binomial logit model for the perceived risk data.

Variable	Level	Latent Class 1	Latent Class 2	Latent Class 3	Latent Class 4
Price	= $40K	.811[c]	.947[c]	3.329[c]	.538[c]
Warranty	= 6yr/60K	-.573[c]	-.899[c]	-.478[c]	-.413[c]
Country of manufacture	= Japan	-1.601[c]	.243[a]	-.521[c]	-.579[c]
Advertising endorsement	= *Road & Track*	-.353[b]	-.129	-.010	-.224[b]
Safety test results	= Approved	-.907[c]	-.409[c]	-.650[c]	-2.200[c]
Newness of brand	= Existing	-1.546[c]	-.044	-.505[c]	-.579[c]
Manufacturing reputation	= Styling	.424[c]	.340[c]	.364[c]	.172[a]
Repair service	= Next day	.376[c]	.015	.112	.242[b]
Loaner car availability	= Yes	-.140	-.074	-.150	-.028
Miles per gallon	= 40	-.371[c]	-.503[c]	-.513[c]	-.396[c]
Special financing	= 1.9% rate	-.085	-.490[c]	-.050	-.183[a]
Braking system	= Anti-lock	-.373[c]	-.194	-.316[c]	-.178[a]
Position in product line	= Luxury	-.358[c]	-.204[a]	-.107	.032
Intercept		-.084	.162	.343	-.184
Mixing proportion		.200	.160	.319	.321

[a] $p<.05$
[b] $p<.01$
[c] $p<.001$

and DeSarbo [15]. However, whereas the stochastic MDS-procedure in [15] focuses on the identification of risk dimensions and individual differences in the weighting of these dimensions, the latent class procedure proposed in this paper focuses on the communalities of the weights of classes of subjects, and tests for the effects of the variables themselves in a logit model of binary choice. Thus, the additional benefit of the latent class procedure over the stochastic MDS procedure arises from capturing consumer heterogeneity with a reduced number of parameters, which facilitates interpretation, and enables significance tests of the effects of stimulus descriptor variables on the perceived risk judgements within segments.

Validation

Five profiles were included in the computerized questionnaire for validation purposes. Each of the 50 respondents rated these 5 profiles on a 10-point scale with endpoints 0 (lowest perceived risk) and 9 (highest perceived risk). These responses were correlated in the [15] study with predicted values formed by the aggregate level PREFPAIRS analysis, the individual level PREFPAIRS analysis, and the stochastic MDS model. The validation correlations were .383, .421, and .458 for these methods, respectively.

We calculate the validation correlations based on the aggregate level (S=1) and S=4 class solutions of the latent class model, as well as those based on a classical two-stage segmentation procedure. In the latter procedure, consumers were classified into 4 classes (using a K-means procedure) on the basis of the coefficients of a logit model estimated for each subject separately. Subsequently, a logit model was estimated within each of the four classes and the coefficients thus obtained were used for prediction of the holdout profiles. The validation correlations were .352 for the $S=1$ class solution, .393 for the two-stage procedure, and .437 for the $S=4$ class solution.

The latent class model clearly outperforms the aggregate level PREFPAIRS and logit models, as well as the two-stage procedure. For the present dataset its performance is somewhat better than the individual level PREFPAIRS analysis, and somewhat less than that of the stochastic MDS model. The higher predictive accuracy of the MDS model accrues at the cost of a much larger number of parameters estimated: whereas the latent class model uses 59 parameters only, the MDS model uses 195. The slightly lower predictive validity of the latent class procedure does not, in our view, offset the increase in interpretability due to the much smaller number of parameters estimated, as compared to the stochastic MDS procedure.

In general, the validation correlations are relatively low. Several explaining hypothesis have been put forward, including respondent fatigue in the paired comparison task [15]. An alternative explanation could be the difference in the types of tasks used in the estimation and validation samples (paired comparisons and ratings, respectively) where the paired comparison task is hypothesized to have a higher reliability. In order to investigate this, we (randomly) eliminated 10 percent of the responses from the estimation sample and used that as a holdout sample. The prediction correlations for these binary responses were .535 for the aggregate and .636 for the 4 class solutions, respectively. This is substantially higher than those for the rating task holdout sample, which supports our hypothesis.

One of the problems with respect to the paired comparisons procedure is the large number (120) of pairs and its possible consequences for respondent fatigue [15]. In order to investigate the sensitivity of our latent class procedure to reducing the number of pairs, we (randomly) eliminated 10, 20, 30, 40, 50, and 60 percent of the pairs and reanalyzed these 6 datasets with our procedure. We used two measures to evaluate performance: the root-mean-squared error of the estimates of the coefficients in the reduced samples versus those estimated in the complete sample (Table 5), and the validation correlation for the ratings of the five holdout profiles.

The root-mean-squared error increased from 10 to 60 percent eliminated pairs (.034, .049, .070, .085, .096, .121), showing that the estimated coefficients increasingly deviate from those estimated from all observations. The validation correlations were .438, .440, .440, .445, .428, and .431 for 10 to 60 percent of the pairs eliminated. Apparently, the validation correlations are not affected up to 40 percent eliminated pairs (and increases somewhat for lower percentages).

DISCUSSION

This paper has presented a method for the external analysis of paired comparison data. The latent class methodology presented in this paper adds to the insights obtained from the analyses by the stochastic MDS procedure proposed by [15] by simultaneously grouping subjects in homogeneous classes and testing the impact of the effects of the components of perceived risk on overall risk within these classes. An elaborate comparison of the internal and external approaches to the analysis of paired comparison data, for example, using a rigorous Monte Carlo experiment is left for further study.

The procedure proposed here is a general procedure for the external analysis of paired comparisons data. The specific application presented in this paper employs a conjoint framework and involves 120 paired comparisons. However, neither the conjoint design, nor the large number of paired comparisons is a requirement for the procedure.

Although our model was developed as a vector model, the extension to ideal point models is simple and straightforward. Moreover, the model can be extended to models that allow for the analysis of a variety of paired comparison experiments, such as models with order effects, models that allow for ties, and ranking models. Finally, future research needs to be conducted in examining algorithm performance via more complete Monte Carlo experimental designs, determining the finite sample properties of the asymptotic standard errors, and investigating alternative measures to utilize for model selection. [Received: August 5, 1992. Accepted: September 30, 1993.]

REFERENCES

[1] Bauer, R. A. Consumer behavior as risk taking. In R. S. Hancock (Ed.), *Proceedings of the 43rd Conference of the American Marketing Association*. Chicago, IL: American Marketing Association, 1984.

[2] *BMDP Manual*. Los Angeles, CA: BMDP Statistical Software Inc., 1990.

[3] Bock, R. D., & Jones, L. V. *The measurement and prediction of judgement and choice*. San Francisco, CA: Holden-Day, 1968.

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

[5] Carroll, J. D. Models and methods for multidimensional analysis of preferential choice (or other dominance) data. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice*. Bern: Hans Huber, 1980, 234-289.

[6] Cochran, W. G., & Cox, G. M. *Experimental design*. New York: John Wiley, 1957.

[7] Cooper, L. G., & Nakanishi, M. Two logit models for external analysis of preferences. *Psychometrika*, 1983, *48*, 607-620.

[8] David, H. A. *The method of paired comparisons*. London: Griffin, 1988.

[9] Dempster, A. P., Laird, N. M., & Rubin P. B. Maximum likelihood from incomplete data via the E-M algorithm. *Journal of the Royal Statistical Society*, 1977, *B39*, 1-38.

[10] De Soete, G. A latent class approach to modelling pairwise preferential choice data. In M. Schader & W. Gaul (Eds.), *Knowledge, data and computer-assisted decisions*. Berlin, Germany: Springer-Verlag, 1990, 103-113.

[11] De Soete, G., & DeSarbo W. S. A latent class probit model for analyzing pick any/*n* data. *Journal of Classification*, 1991, *8*, 45-63.

[12] Dillon, W. R., Kumar, A., & Smith de Borrero, M. Capturing individual differences in paired comparisons: An extended BTL model incorporating descriptor variables. *Journal of Marketing Research*, 1993, *30*, 42-51.

[13] Dowling, G. R. Perceived risk: The concept and its measurement. *Psychology and Marketing*, 1986, *3*, 193-210.

[14] Fechner, G. T. *Elements der psychophysik*. Leipzig, Germany: Breitkopf and Haertel, 1860.

[15] Havlena, W. J., & DeSarbo, W. S. On the measurement of perceived risk. *Decision Sciences*, 1991, 22, 927-939.

[16] Jacoby, J., & Kaplan, L. B. The components of perceived risk. In M. Venkatesan (Ed.), *Proceedings of the 3rd Annual Conference of the Association for Consumer Research*. Chicago, IL: Association for Consumer Research, 1972.

[17] Kamakura, W. Estimating flexible distributions of ideal-points with external analysis of preferences. *Psychometrika*, 1991, *56*, 419-431.
[18] Kamakura, W., & Russell, G. J. A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 1989, *26*, 379-390.
[19] McLachlan, G. J., & Basford, K. E. *Mixture models: Inference and application to clusting*. New York: Marcel Dekker, 1988.
[20] Peter, J. P., & Ryan, M. J. An investigation of perceived risk at the brand level. *Journal of Marketing Research*, 1976, *13*, 184-188.
[21] Ross, I. Perceived risk and consumer behavior. A critical review. In M. J. Schlinger (Ed.), *Advances in consumer research* (Vol. 2). Urbana, IL: Association for Consumer Research, 1976.
[22] Stem, D. E., Jr., Lamb, C. W., Jr., & MacLachlan, D. L. Perceived risk: A synthesis. *European Journal of Marketing*, 1977, *11*, 312-319.
[23] Thurstone, L. L. A law of comparative judgement. *Psychological Review*, 1927, *34*, 273-286.
[24] Wedel, M., & DeSarbo, W. S. A latent class binomial logit methodology for the analysis of paired comparison data. Memorandum from the Institute of Economic Research, nr. 467, Faculty of Economics, University of Groningen, Groningen, Netherlands, 1992.
[25] Wedel, M., & Kistemaker, C. K. Consumer benefit segmentation using clusterwise linear regression. *International Journal of Research in Marketing*, *6*, 45-59.

Michel Wedel is Professor of Marketing Research and Market Structure Analysis at the Department of Marketing and Marketing Research, Faculty of Economics, University of Groningen in the Netherlands. He received a degree in biometrics at the University of Leiden, and his Ph.D. in marketing at the University of Wageningen. He has published over 60 articles in the fields of human nutrition and epidemiology. His major marketing publications have appeared in journals such as *Journal of Retailing, Journal of Business Research, International Journal for Research in Marketing,* and *Journal of Marketing Research*. His main current research interests are in classification, multidimensional scaling, conjoint analysis, market segmentation, and market structure analysis.

Wayne S. DeSarbo is Sebastian S. Kresge Distinguished Professor of Marketing and Statistics at the Graduate School of Business, the University of Michigan. He received his B.S. degree in economics from the Wharton School of the University of Pennsylvania, his M.A. degrees in sociology, administrative science/operations research, and marketing from Yale University and the University of Pennsylvania (UP), and his Ph.D. in operations research and econometrics from UP. He has published over 80 articles in journals such as *Journal of Marketing Research, Journal of Consumer Research, Journal of Mathematical Psychology, Marketing Science, Journal of Classification*, and others. His methodological interests lie in multidimensional scaling, classification and multivariate statistics, specifically substantive marketing problems in positioning, market structure, consumer choice, market segmentation, and competitive strategy.