A comparison of two-stage segmentation methods for choice-based conjoint data: a simulation study

Marjolein Crabbe, Bradley Jones and Martina Vandebroek

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

A Comparison of Two-Stage Segmentation Methods for Choice-Based Conjoint Data: A Simulation Study

Marjolein Crabbe
Bradley Jones
Martina Vandebroek

Abstract

Due to the increasing interest in market segmentation in modern marketing research, several methods for dealing with consumer heterogeneity and for revealing market segments have been described in the literature. In this study, the authors compare eight two-stage segmentation methods that aim to uncover consumer segments by classifying subject-specific indicator values. Four different indicators are used as a segmentation basis. The *forces*, which are subject-aggregated gradient values of the likelihood function, and the *dfbetas*, an outlier detection measure, are two indicators that express a subject's effect on the estimation of the aggregate partworths in the conditional logit model. Although the conditional logit model is generally estimated at the aggregate level, this research obtains *individual-level partworth estimates* for segmentation purposes. The respondents' raw *choices* are the final indicator values. The authors classify the indicators by means of cluster analysis and latent class models. The goal of the study is to compare the segmentation performance of the methods with respect to their success rate, membership recovery and segment mean parameter recovery. With regard to the *individual-level estimates*, the authors obtain poor segmentation results both with cluster and latent class analysis. The cluster methods based on the *forces*, the *dfbetas* and the *choices* yield good and similar results. Classification of the *forces* and the *dfbetas* deteriorates with the use of latent class analysis, whereas latent class modeling of the *choices* outperforms its cluster counterpart.

Keywords: two-stage segmentation methods, choice-based conjoint analysis, conditional logit model, market segmentation, latent class analysis

1   Introduction

Since Smith's (1956) leading article, market segmentation has steadily grown to one of the most important aspects of modern marketing and marketing research. The ongoing interest in market segmentation is fed by the presence

of heterogeneity on both the supply and the demand side of the market. The heterogeneity in buyers' preferences and needs, which ultimately results in different product choices, splits the market into distinct, more or less homogeneous segments of consumers. Market segmentation and its resulting actions such as product modification or target offering and pricing, tailoring marketing strategies to one specific target segment, have therefore been main objectives in strategic management for a long time (Johnson 1971; Biggadike 1981). In order to increase sales and consequently profits, companies can either modify their product so that it appeals to the majority of the consumers or one can opt to address one single segment.

The list of possible market partitions is endless, therefore a reliable segmentation basis is crucial. Haley (1968), Wind (1978) and Currim (1981) argued that benefit segmentation is most powerful in search for market segments. The idea behind benefit segmentation is that the presence of among-person variation in the benefits they are seeking is believed to be the true cause for discrepancies in buyer behavior and the existence of niches in the market. The utility or benefit a consumer receives from a product or service results from the product attributes. The utilities from the product features are therefore used in benefit segmentation to identify groups with similar preferences. Benefits sought as segmentation basis is preferred over a priori observed segmentation bases such as geographic or demographic characteristics. Though, dissimilarities in these person variables over segments can sometimes give a deeper understanding of the segments found and can unfold a profile of an average individual in each segment (Haley 1968; Green and Krieger 1991; Gupta and Chintagunta 1994).

One of the most widely-used methods to examine the preference structure of consumers is conjoint analysis (Green and Srinivasan 1978; 1990; Green and Krieger 1991), where respondents are asked to evaluate potential bundles of attributes, i.e. product profiles. Multiple features of the products are thus considered jointly, hence conjoint analysis. One way is to show the selected profiles one at a time and ask respondents to rate their preference or likelihood of purchase on a given measurement scale. More recently, choice-based conjoint analysis has grown in popularity. In multifactor choice situations participants must choose one alternative from a pair (paired comparison) or a set of profiles. By modeling the ratings-based preferences or the choices with appropriate conjoint models, the valuation and relative importance of each product attribute is analysed.

Elrod et al. (1992) and Karniouchina et al. (2009) observed no systematic differences between the parameter estimates in ratings-based (RB) and choice-based (CB) conjoint models. Nevertheless, CB conjoint analysis is favored by most researchers because of the resemblance with real-life purchase decisions. Choosing a preferred product from a set of alternatives is what consumers actually do in a marketplace. Although one can get easily information overload in CB conjoint analysis, especially when profiles with more than six attributes need to be compared, a choice task is assumed to be simple, natural and easier to perform than ratings (Desarbo et al. 1995; Sawtooth Software 2008). On the other hand, CB data only provide information about the actual choice and not

about the strength of preference.

Besides their specific drawbacks, both RB and CB conjoint analysis have some common additional disadvantages. A poor choice of profiles, a non-representative selection of participants and the reliance on hypothetical products may lead to dubious results (Alriksson and Öberg 2008). Nevertheless, conjoint analysis is frequently implemented in diverse fields. For example, not only for making various marketing decisions such as pricing, product identification and advertising (Louviere and Woodworth 1983; Alriksson and Öberg 2008; Song et al. 2009; Chaim et al. 2009) but also for decision making in the health sector (Lancsar and Louviere 2008) conjoint analysis is extensively applied. But its major commercial application is market segmentation (Cattin and Wittink 1982; Wittink and Cattin 1989; Wittink et al. 1994). Besides numerous examples for consumer good companies, conjoint analysis also knows many successful segmentation results in health (Cunningham et al. 2009), environmental research (Bigsby and Ozanne 2002) and in the transportation field (Greene and Hensher 2003).

Traditionally, segmentation has been done with a two-stage segmentation method combining ratings-based conjoint models and cluster analysis (Currim 1981; Hagerty 1985; Vriens et al. 1996). After preference models are estimated on the individual level, the model parameters are grouped with a non-overlapping cluster algorithm. In this paper this cluster-based segmentation is reviewed for choice-based conjoint data and compared with alternative methods. The conditional logit (CL) model, which is used to model the CB data, is generally estimated at the aggregate level using the entire sample of respondents. In order to uncover segments in the data, the model partworths are estimated at the individual-level and clustered afterwards. Subjects with a similar preference structure will have similar estimates and form homogeneous segments to the extent that the estimated attribute utilities are accurate (Currim 1981; Kamakura 1988). Unfortunately estimation error is unavoidable and possible bias in the *individual-level parameter estimates* is ignored, which is the main drawback of such cluster-based segmentation. Besides clustering the *individual-level partworth estimates*, Green and Srinivasan (1978) and Vriens et al. (1996) therefore mention the possibility to cluster the profile ratings instead. However, they found that the estimates of the model parameters are more discriminating than the ratings themselves. Extending the idea of Green and Srinivasan and Vriens et al. to choice-based conjoint data, we will cluster the *choices* made by the respondents in each choice set in search of market segments.

Further, two more sets of subject-specific indicator values are introduced in this paper for segmentation purposes. These additional indicators are the *forces*, which are subject-aggregated gradient values of the likelihood function, and the *dfbeta values*, a conditional logit model extension of the outlier detection measure in linear regression analysis. Both sets of continuous values are derived from the aggregate model parameter estimates in the conditional logit model. The idea of using the *forces* to detect consumer segments was introduced recently in the statistical program JMP (SAS 2009). The potential segmentation ability of the *forces* is based on the fact that they express the force a subject

is performing on the aggregate estimates of the model parameters in the conditional logit model. The more outlying a subject is with respect to some or all model partworths, the larger its corresponding *forces* will be. Consumers with similar preferences influence the estimates similarly and are therefore homogeneous with respect to the *forces*. The *dfbetas* of a linear regression model also express how subjects affect parameter estimation. Since this effect could potentially say something about the similarity in preferences, the *dfbetas* are extended to the conditional logit model and also used as segmentation basis.

Despite the common use of cluster analysis, Hagerty (1985) points out that the assumption of non-overlapping clusters is not always realistic. Therefore the aforementioned indicators and thus the corresponding consumers will also be classified into segments by means of latent class analysis. Wolfe (1970) was the first to associate latent class models with cluster analysis. Latent class analysis is a model-based cluster approach where cases are classified into clusters according to posterior probabilities of class membership. Rather than classifying subjects into clusters based on some arbitrary "similarity" measure, latent class analysis takes the uncertainty of class membership into account. Both for categorical as for continuous indicators, latent class models are elaborately discussed in the literature (Goodman 1974; Hagenaars and McCutcheon 2002; Magidson and Vermunt 2002; Vermunt and Magidson 2002; Magidson and Vermunt 2004).

The four sets of indicator values of the subjects, i.e. *the individual-level partworth estimates*, the *choices*, the *forces* and the *dfbetas*, and the two classification methods, i.e. cluster and latent class analysis, define eight segmentation methods. One should bear in mind that these two-stage methods are relatively straightforward and that more advanced one-stage methods, which simultaneously perform market segmentation and conjoint analysis, exist to deal with preference heterogeneity. As such, finite mixtures of conjoint models are discussed in the literature. Desarbo et al. (1992) and Wedel and Desarbo (1995) introduce a latent class approach for metric conjoint analysis whereas Desarbo et al. (1995) apply finite mixtures to the CL model. An extension of the conditional logit mixture model is the use of covariates to predict class membership (Gupta and Chintagunta 1994; Vermunt and Magidson 2005b). Note that, although these one-stage methods and the two-stage methods comprising latent class modeling of the indicators both make use of latent class analysis, only with the one-stage methods the segment-specific partworths are estimated. In contrast, for the two-stage methods the latent class cluster model is applied to indicator values and only results in the classification of consumers into segments. An additional step is necessary to obtain segment-specific model partworth estimates for each segment found.

The heterogeneity in model partworths across respondents can also be accounted for with mixed logit or random effects models (Allenby and Rossi 1999; Greene and Hensher 2003; Colombo et al. 2008). In these models the parameters are assumed to follow a multivariate, most frequently normal, distribution. Instead of latent classes, the random effects explain preference heterogeneity. Allenby et al. (2005) and Orme and Howell (2009) argument that also in mixed

4

logit models the use of well-chosen covariates can improve the parameter estimates and therefore the identification of the heterogeneity.

In case of a considerable amount of within-segment heterogeneity in the model partworths, the finite mixture approach is insufficient since it may lead to huge overestimation of the number of classes. In the presence of substantial subpopulations on the other hand, the random effects model may also be inadequate. Therefore Lenk and Desarbo (2000) introduce a model that considers both finite mixtures of generalized linear models and random effects within mixtures. The distribution of the model coefficients is then assumed to be a mixture of, most frequently normal, distributions. In this way the flexibility of the latent class model and the parsimony of the mixed logit model are incorporated.

The substantial complexity of the more advanced models and the need for more sophisticated software in order to estimate them, increases the interest in the segmentation ability of more straightforward methods. In this light the focus of this research is on the segmentation performance of the eight two-stage methods introduced earlier, which will be compared by means of a simulation study. In order to get a first impression of the segmentation power of the indicators used and the classification methods applied, an overview of some relevant research is given in Table 1. Some of these results are described in more detail below.

Currim (1981) used cluster techniques to classify respondents by means of their individual-level partworth estimates of RB models. He estimated a segment-based choice model with the segments found and observed considerable predictive advantages in comparison to the aggregate model. However, many authors point out that a non-overlapping cluster solution for the individual-level partworths of rating models does not correctly describe respondents' preference heterogeneity. For choice-based conjoint models, the classification of the individual-level partworth estimates was done by Cunningham et al. (2009). They used latent class models and could successfully classify the respondents into interpretable segments. Note that they obtained the estimates of the utility coefficients at the individual level with Hierarchical Bayes (HB) techniques. We investigate whether these results hold when the individual-level estimates are obtained with ordinary maximum likelihood.

Bigsby and Ozanne (2002) clustered product attribute ratings for wood outdoor furniture and found interpretable consumer segments. Desarbo et al. (1995) extended the idea to CB conjoint analysis. They clustered observed choice frequencies for households in the data and compared these results with the performance of a finite mixture of choice models, for which an interpretable four-class solution was found. Classifying the households into four segments with cluster analysis gave partworth estimates with less variation between segments. Moreover, the segment-level estimates were even similar to the aggregate estimates. Also Desarbo et al. (1992) and Vriens et al. (1996) obtained superior segmentation results for finite mixtures of conjoint models in comparison to traditional cluster techniques. This brief enumeration shows that various, sometimes contradicting, results are found in the literature.

The remainder of this paper is organized as follows. In the next section, the

5

conditional logit model and the two-stage segmentation methods are described. We then present the design of the simulation study, followed by the results. The last section includes some discussion and conclusions.

2 Methodology

2.1 Conditional Logit Model

The results from the choice-based conjoint experiments are analyzed with a conditional logit model (Train 2003). The utility that decision maker n receives from alternative j in choice set s is given by a deterministic and a random part.

$$U_{njs} = V_{njs} + \varepsilon_{njs} \tag{1}$$
$$= \mathbf{x}'_{njs}\boldsymbol{\beta} + \varepsilon_{njs} \tag{2}$$

The deterministic part V_{njs} of the utility function is determined by the observed attributes of the alternative, represented by the K-dimensional vector $\mathbf{x}_{njs}$. The sensitivity of the utility to the level of the attributes is expressed by the model partworths, denoted by the K-dimensional vector $\boldsymbol{\beta}$. The partworths express the importance of each attribute for the respondents. The effect on the utility of factors not included in the model is captured by ε_{njs}.

The CL model assumes that each ε_{njs} is independently and identically extreme value distributed. Therefore a closed form solution of the probability that decision maker n chooses alternative j in choice set s can be found,

$$p_{njs} = \frac{e^{\mathbf{x}'_{njs}\boldsymbol{\beta}}}{\sum_{t=1}^{J} e^{\mathbf{x}'_{nts}\boldsymbol{\beta}}}. \tag{3}$$

To estimate the model partworths, maximum likelihood techniques are applied. The log-likelihood function for the CL model is maximized

$$LL(\boldsymbol{\beta}) = \sum_{n=1}^{N} \sum_{j=1}^{J} \sum_{s=1}^{S} y_{njs} \ln(p_{njs}(\boldsymbol{\beta})), \tag{4}$$

where y_{njs} is 1 when decision maker n chooses alternative j in choice set s and 0 otherwise. The CB data, i.e. the choices made by each respondent and maximum likelihood estimates of the model parameters, are now used to uncover segments of consumers with similar preferences and thus similar choice behavior.

2.2 Two-stage Segmentation Methods

Each segmentation method embodies two stages and starts from a set of indicator values, generated for each subject in the sample. By means of cluster analysis or latent class models these indicators and corresponding subjects are classified into clusters or latent classes, which represent the segments. Four indicators are introduced i.e. the *forces*, the *individual-level estimates* of the model partworths, the *dfbetas* and the *choices* themselves. In combination with two classification methods we obtain eight segmentation procedures.

Table 1: Overview of the literature on two-stage segmentation methods

Authors	Indicator	Type of data	Classification	Conclusion
Green and Srinivasan (1978)	Preference ratings of the profiles	RB	Cluster	· Attribute utilities are more discriminating than preferences
Currim (1981)	Individual-level partworth estimates with OLS	RB	Cluster	· Segment-based choice model outpredicts aggregate model
Hagerty (1985)	Individual-level partworth estimates with OLS	RB	Cluster	· Optimal weighting procedure outperforms traditional non-overlapping cluster analysis
Kamakura (1988)	Individual-level partworth estimates with OLS	RB	Cluster	· Maximizing predictive validity of segment-level estimates outperforms cluster analysis
Desarbo et al. (1992)	Individual-level partworth estimates with OLS	RB	Cluster	· Finite mixture of conjoint models gives interpretable latent class solution for case study · Different segment solutions for different cluster methods
Desarbo et al. (1995)	Observed choice frequencies of households	CB	Cluster	· Finite mixture of conjoint models gives interpretable latent class solution for case study · Less variation in segment results obtained with cluster analysis
Vriens et al. (1996)	Individual-level partworth estimates with OLS and preference ratings of the profiles	RB	Cluster	· Finite mixture of conjoint models outperforms two-stage segmentation methods · Two-stage method clustering the individual-level estimates outperforms cluster analysis of the ratings
Bigsby and Ozanne (2002)	Product attribute ratings	RB	Cluster	· Interpretable segments found in case study with cluster analysis of ratings
Cunningham et al. (2009)	Individual-level partworth estimates with HB	CB	Latent Class	· Interpretable latent class solution found in case study with better fit than aggregate analysis

2.2.1 Indicators

Forces
An iterative Newton-Raphson step to obtain the maximum likelihood estimates of the vector of model parameters $\boldsymbol{\beta}$ equals $\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + (-\mathbf{I}_t^{-1})\boldsymbol{g}_t$ (Train 2003). At the estimates $\Delta = \mathbf{I}^{-1}\boldsymbol{g} = 0$ since the total gradient $\boldsymbol{g}$ of the log-likelihood evaluated at the estimates equals zero. The matrix $\mathbf{I}^{-1}$ is the inverse of the Fisher information matrix, which consists of the negatives of the second order derivatives of the log-likelihood function. The information matrix is the inverse of the variance-covariance matrix of the parameter estimators. For the CL model the Fisher information matrix equals (see for instance Kessels et al. 2006)

$$\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}) = \sum_{n=1}^{N} \sum_{s=1}^{S} \mathbf{X}'_{ns}(\mathbf{P}_{ns} - \mathbf{p}_{ns}\mathbf{p}'_{ns})\mathbf{X}_{ns}, \tag{5}$$

where $\mathbf{X}_{ns} = [\mathbf{x}'_{n1s}, ..., \mathbf{x}'_{nJs}]'$, $\mathbf{p}_{ns} = [p_{n1s}, ..., p_{nJs}]'$, $\mathbf{P}_{ns} = \mathrm{diag}[p_{n1s}, ..., p_{nJs}]$. By means of the expression of the log-likelihood function in (4) one finds that

$$\Delta \quad = \quad \mathbf{I}_{\hat{\boldsymbol{\beta}}}^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} LL(\hat{\boldsymbol{\beta}}) \tag{6}$$

$$= \quad \sum_{n=1}^{N} \sum_{j=1}^{J} \sum_{s=1}^{S} \mathbf{I}_{\hat{\boldsymbol{\beta}}}^{-1} \, y_{njs} \, \frac{\partial}{\partial \boldsymbol{\beta}} \ln(p_{njs}(\hat{\boldsymbol{\beta}})) \tag{7}$$

$$= \quad \sum_{n=1}^{N} \sum_{j=1}^{J} \sum_{s=1}^{S} \underbrace{\mathbf{I}_{\hat{\boldsymbol{\beta}}}^{-1} \, y_{njs} \left[\mathbf{x}'_{njs} - \frac{\sum_{t=1}^{J} \mathbf{x}'_{nts} e^{\mathbf{x}'_{nts}\hat{\boldsymbol{\beta}}}}{\sum_{t=1}^{J} e^{\mathbf{x}'_{nts}\hat{\boldsymbol{\beta}}}} \right]}_{\Delta_{njs}}, \tag{8}$$

with $\hat{\boldsymbol{\beta}}$ the aggregate partworth estimates. The *forces* for decision maker n now are defined as (SAS 2009)

$$\bar{\Delta}_n = \sum_{j=1}^{J} \sum_{s=1}^{S} \frac{\Delta_{njs}}{J \times S}, \tag{9}$$

with $J \times S$ the number of profiles in the design. By construction, the subject *forces* sum up to zero over the respondents in the data set. They express the force a subject is practising on the aggregate maximum likelihood estimates of the model parameters. Classifying these *forces* into clusters or latent classes could reveal segments in the data, since subjects with similar choice behavior will also be homogeneous in terms of influence on the estimates.

Individual-level Estimates of the Partworths
The CL model assumes a single set of model parameters. The model is fit at the aggregate level and partworths are estimated for the entire sample of respondents. When enough choice sets are included in the design, maximum likelihood estimates of the model parameters can be obtained for each individual. By estimating the partworths at the individual level, one can account for

the preference heterogeneity present in the data since similar preference structures should result in similar *individual-level estimates*. Note that the mixed logit model also assumes individual-specific partworths. In contrast, the mixed logit model assumes a multivariate distribution for the model parameters and the distribution parameters are estimated.

Dfbetas

The *dfbetas* were introduced in linear regression as outlier detection measures. They measure a subject's influence on the estimation of a model parameter. The *dfbeta* value for the kth model parameter and decision maker n equals (Lattin et al. 2003)

$$dfbetas_k = \frac{\hat{\beta}_k - \hat{\beta}_k(n)}{s(\hat{\beta}_k)(n)}, \tag{10}$$

with $\hat{\beta}_k$ the estimate of the kth model parameter for the whole sample, $\hat{\beta}_k(n)$ the estimate with respondent n left out of the sample and $s(\hat{\beta}_k)(n)$ the estimate of the standard deviation for the kth parameter's estimate again leaving subject n out of the sample. A $dfbetas_k$ value equal to zero means that subject n has no influence on the estimation of parameter β_k. The larger the influence measure, the larger the evidence that one deals with an outlier. For segmentation purposes, this concept is generalized to the conditional logit model. As with the *forces*, one expects that similar choice behavior will influence parameter estimation similarly. For the CL $dfbetas_k$, the denominator now equals $\sqrt{[(\mathbf{I}(\mathbf{X}, \hat{\boldsymbol{\beta}})(n))^{-1}]_{kk}}$ with $\mathbf{I}(\mathbf{X}, \hat{\boldsymbol{\beta}})(n)$ the Fisher information matrix in (5) computed discarding subject n.

Choices

The final indicator values are the raw respondents' *choices*. In contrast to the aforementioned indicators, they do not make use of any model or estimation method.

Note the discrepancies in dimensions between the different indicators. The dimension of the respondents' *choices* equals the number of choice sets in the design. For the remaining indicators, the *forces*, the *individual-level estimates* and the *dfbetas*, the dimension is equal to the number of parameters in the model. The dimension of the indicator will be denoted by T.

2.2.2 Classification Methods

Once these indicators are obtained, the second stage of each segmentation method considered aims at classifying these values with either cluster analysis or by means of latent class models. Everitt et al. (2001) describe cluster analysis as a meaningful classification of data for which the objects within a cluster are similar but different from the members of other groups with respect to some cluster variables. One seeks for homogeneity within and heterogeneity between clusters. Latent class analysis is a model-based clustering approach

9

(Vermunt and Magidson 2002), since a statistical model is hypothesized for the population from which the sample at hand is drawn. In contrast to standard cluster analysis which deals with non-overlapping clusters, uncertainty about an object's class membership is taken into account in latent class analysis.

Cluster Analysis
The indicators and corresponding subjects are classified into clusters combining both a hierarchical and a non-hierarchical cluster approach. Hierarchical clustering starts from N clusters i.e. each cluster comprises a subject. Merging step by step the two most similar clusters, one cluster consisting of all subjects is left at the end. To determine the appropriate amount of clusters for the data at hand a number of evaluation statistics are available (Sharma 1996). In non-hierarchical clustering the number of clusters is fixed beforehand. After selecting a specific set of observations as the initial seeds or the initial cluster means, subjects are assigned to the cluster with the nearest cluster mean. Cluster means are recalculated and subjects are reassigned until no improvement can be obtained with respect to some convergence criterion. In a comparison study Punj and Stewart (1983) found that Ward's minimum variance method, wherein clusters are merged by minimizing the total within-group or within-cluster sum of squares, outperforms the other hierarchical approaches for most data. They also concluded that non-hierarchical methods outperform the hierarchical procedures if non-random starting points, e.g. cluster means obtained from hierarchical analysis, are chosen.

In this paper non-hierarchical clustering is used to construct the segments. The initial seeds, for the number of clusters considered, are obtained with Ward's method. The non-hierarchical cluster solutions are then evaluated by means of their R-Squared value and the ratio of the within-cluster standard deviation on the total standard deviation, in order to choose the appropriate number of clusters for the data present. Also the Pseudo F Statistic is taken into account for this, all by all, subjective decision.

Latent Class Analysis
In addition to cluster analysis, latent class models can be used for the classification of the indicators. The basic structure for a latent class cluster model is (Vermunt and Magidson 2005a)

$$f(\mathbf{y}_n) = \sum_{g=1}^{G} \pi(g) f(\mathbf{y}_n|g), \tag{11}$$

with g a single nominal latent variable, $\mathbf{y}_n$ the T indicator values and $\pi(g)$ the estimated prior probability of belonging to latent class g. This prior probability also represents the size of class g. Note that no covariates are included in the model.

In case of categorical indicators, e.g. the respondents' *choices*, the standard

latent class model is obtained

$$\pi(\mathbf{y}_n) = \sum_{g=1}^{G} \pi(g) \prod_{t=1}^{T} \pi(y_{nt} = m_t | g), \tag{12}$$

with restrictions

$$\sum_g \pi(g) \quad = \quad 1 \tag{13}$$

$$\sum_{m_t} \pi(y_{nt} = m_t | g) \quad = \quad 1 \qquad \forall t. \tag{14}$$

Given the membership of a subject in a latent class, the indicator values are assumed to be independent of one another. This assumption is referred to as local independency.

For continuous indicators, e.g. the *forces*, the *individual-level partworth estimates* and the *dfbetas*, the least restrictive form of the latent class cluster model in (11) allows for local dependencies between all indicator values, more specifically it is assumed that the $\mathbf{y}$'s come from latent class-specific multivariate normal distributions.

Based on the prior probabilities, the posterior probabilities of class membership can be calculated

$$\hat{\pi}(g|\mathbf{y}_n) = \frac{\hat{\pi}(g)\hat{f}(\mathbf{y}_n|g)}{\hat{f}(\mathbf{y}_n)}. \tag{15}$$

Each object is classified in the latent class with the highest posterior probability. For technical details about the estimation procedure one is referred to Vermunt and Magidson (2005a). To decide on the appropriate number of latent classes for the data at hand the BIC information criterion is minimized i.e. $BIC_{LL} = -2LL + (\ln N)K$ (Vermunt and Magidson 2005a), where N and K respectively equal the number of subjects in the data set and the number of parameters in the model.

Note that although preference heterogeneity will be assumed and simulated in the data, latent class analysis typically models no within-component variation. The presence of significant preference variation in the segments could therefore deteriorate the segmentation performance of the latent class methods.

3   Design of the Simulation Study

We employ a simulation study to give insight about the performance of the different segmentation methods described. Besides the effect of the indicators that are used for segmentation and the classification method applied, the potential effects of some additional factors are taken into account in the simulation setup. The number of segments, the number of respondents, the within-segment heterogeneity and the mean separation between the within-segment partworth

11

means in the simulations can influence the segmentation results. These additional factors and their different settings are given in Table 2 and described below. As not all combinations are feasible (e.g. with only one segment, separation between segments is not meaningful), the design is not a full factorial. In case of one segment there are only 6 factor level combinations. For two or three segments there are another $2^3 \times 3 = 24$ experimental conditions. To enable generalization of the results, we considered 10 repetitions for each experimental condition. This results in 300 data sets that are analyzed using the eight segmentation methods.

Table 2: Factors included in the study

Factor	Levels
A. Number of segments	1, 2, 3
B. Number of respondents	100, 300
C. Homogeneous versus diffuse segments (variance)	0, 0.05, 0.10
D. Mean separation between segment means	1, 2

3.1 Data Generation

The data sets consist of individual simulated choices for all the choice sets in the design. The design used in the simulation study is a nearly orthogonal full profiles (Chrzan and Orme 2000) design and is given in Appendix A. It includes 16 choice sets with 3 alternatives. Each alternative is defined by three attributes, which in turn have three levels each ($3^3/3/16$). It is assumed that all subjects evaluate the same design. The dimension of the *choices* indicator thus equals 16 in this study since there are 16 choice sets included in the design. Effects coding for the three three-leveled attributes results in 6 model partworths. So, for the *forces*, the *individual-level estimates* and the *dfbetas* the dimension is 6.

For the construction of the choices individual-level partworths $\boldsymbol{\beta}_n$ are simulated, based on the ideas of Vriens et al. (1996) and Andrews and Currim (2003). First within-segment mean partworths $\boldsymbol{\beta}_g$, with β_{gk} the mean value for the kth partworth in the gth segment, are simulated. The elements of the vector of mean parameter values $\boldsymbol{\beta}_1$ for the first segment are drawn from the uniform distribution $\mathcal{U}[-1.7,1.7]$. Next a vector of separations $\boldsymbol{d}$ is randomly generated, as is a vector of signs S for $\boldsymbol{d}$. The elements of the vector $\boldsymbol{d}$ are sampled from a normal distribution for which the standard deviation equals 10% of the mean. The mean partworths for the second and, if present, the third segment now respectively equal $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_1 + S\boldsymbol{d}$ and $\boldsymbol{\beta}_3 = \boldsymbol{\beta}_1 - S\boldsymbol{d}$. To account for within-segment heterogeneity individual-specific values are added to the mean model parameters. A $N \times K$ matrix of random values δ_{nk} is considered, with N and K respectively the number of subjects and the number of model parameters (6). The values of the matrix are drawn from a normal with mean 0 and variance σ^2. By considering several values for σ^2 either homogeneous or diffuse segments are obtained. The subject-specific vectors $\boldsymbol{\delta}_n$ are randomly assigned to the different segments and the true individual-level partworths are computed

12

as $\beta_{nk(g)} = \beta_{gk} + \delta_{nk}$. By means of these individual-level partworths the choices for all choice sets are simulated for each respondent.

3.2   Factors Included in the Study

3.2.1   Number of Segments

As in Andrews et al. (2002a,b), data from one, two or three segments are simulated. Similar numbers of segments are used by Hagerty (1985), Wedel and Desarbo (1995) and Lenk and Desarbo (2000). In case of two or three segments, equally large segments are assumed. As in Vriens et al. (1996), there is no reason to assume that any segmentation method would benefit from either equal or unequal segment sizes. For the latent class methods, one expects to correctly estimate the number of segments more often when fewer segments are simulated since a larger number of segments requires the estimation of more class memberships and latent class-specific model parameters which could decrease the performance of the estimation procedures and thus the accuracy of the decision. Also for the segmentation methods using cluster analysis, an increase in the number of segments is expected to deteriorate their segmentation performance.

3.2.2   Number of Respondents

Based on the number of respondents used in Vriens et al. (1996), Andrews et al. (2002a,b) and Andrews and Currim (2003), the number is set equal to either 100 or 300. Similar values are found in Hagerty (1985), Wittink et al. (1994), Lenk and Desarbo (2000) and Greene and Hensher (2003). One might expect that a minimum of 100 subjects is sufficient to discover potential clusters in the data, but as Dolnicar (2002) states there is no rule-of-thumb for choosing the right sample size in cluster analysis. Further it is not quite clear whether an increase in the number of subjects from 100 to 300 will significantly improve the cluster segmentation results. In latent class analysis, also an increase in the number of respondents demands the estimation of more segment membership parameters which could adversely affect the estimation procedures' performance.

3.2.3   Homogeneous versus Diffuse Segments

In order to evaluate the effect of within-segment heterogeneity on the segmentation results, the level of the within-component variance σ^2 is altered. Similar to Hagerty (1985), Vriens et al. (1996) and Andrews et al. (2002b) the values for this factor are set equal to 0.05 and 0.10. Since the latent class segmentation methods assume no within-component heterogeneity a third level of zero variance is added. In cluster analysis non-overlapping clusters are assumed, thus larger within-segment variation will not facilitate segmentation. But also for latent class analysis, where uncertainty about class membership is taken into account, an increase in within-component heterogeneity is believed to deteriorate the performance of the segmentation methods.

3.2.4   Mean Separation between Segment Means

In case of more than one segment, Vriens et al. (1996) and Andrews et al. (2002b), dealing with ratings data, multiplied the simulated segment mean partworths by a factor two to increase the dissimilarity between segments. It was pointed out in Andrews et al. (2002a) that such a manipulation of the segment separation only changes the scale factor of the CL model. Andrews and Currim (2003) came up with the idea of altering the mean separation between the segment means of the partworths instead, i.e. the mean of the normal distribution from which the elements of the separation vector d are drawn. A mean separation of 1 is considered to create similar segments, whereas dissimilar segments correspond to a mean separation of 2. Obviously, for all segmentation methods it is expected that an increase in the mean separation between segment means will improve the segmentation results.

In order to visualize the effect of the different settings of the factors on the simulated individual-level partworths, principal component analysis is performed on these simulated values. Scatterplots of the scores for the first two principal components are given in Appendix B.

3.3   Measures of Performance

The main objective of the study is the comparison of the segmentation methods with respect to their ability to uncover the true number of segments in the simulated data sets and the ability to classify the respondents in the correct segment. As mentioned earlier, some evaluation statistics and the BIC information criterion are used in the cluster and the latent class methods respectively to choose the number of segments in each data set. The performance of the segmentation methods is then measured by their success rate, which is the percentage of cases in which the true number of simulated segments was detected. This percentage of correctly estimated numbers of segments (%CORRNUM) will be compared across segmentation methods and across the levels of the additional factors considered in the simulation setup. An erroneous choice of segments either results in underfitting or in overfitting. Since the presence of small segments in case of overfitting could produce large and unstable parameter values and thus severe bias, Andrews and Currim (2003) prefer underfitting to an overestimation of the number of segments. On the other hand Nylund et al. (2007) argue that the true g class solution could be extracted from a $g + 1$ solution, in case one of the classes is very small and hard to identify or isn't meaningful. Therefore underidentifying the number of classes can be considered worse than overestimating it.

Besides comparing the success rates of the segmentation methods, one can also compare their performance in terms of membership and coefficient recovery. For this purpose we set the number of clusters or latent classes equal to the true number of simulated segments. Under this fixed and correct number, subjects are classified and for each cluster or latent class separately the maximum like-

lihood estimates of the model parameters are obtained. On the one hand, each respondent belongs to its true segment g_n, with mean partworths $\boldsymbol{\beta}_{g_n}$. On the other hand, the subject belongs to the cluster or latent class $\hat{g}_n$ it is classified in, with aggregate maximum likelihood estimates $\hat{\boldsymbol{\beta}}_{\hat{g}_n}$. To verify whether the segmentation methods correctly classify the majority of the subjects and whether the obtained maximum likelihood estimates sufficiently approximate the true mean partworths of the segments, two additional measures of performance are introduced, based on the measures given in Vriens et al. (1996) and Andrews et al. (2002a,b)

- the percentage of correctly classified subjects (%CORRCLASS), which is the percentage of subjects that are classified into their true segment

- the root-mean-squared-error between the true and estimated values of the mean partworths in each segment

$$\text{RMSE}(\boldsymbol{\beta}) = \sqrt{\sum_{n=1}^{N} \sum_{k=1}^{K} \frac{(\beta_{g_n k} - \hat{\beta}_{\hat{g}_n k})^2}{N \times K}}.$$

4 Results of the Simulation Study

ANOVAs were used to analyze the effect of the indicators used and the classification method applied on the performance of the segmentation methods with respect to the three performance measures. Both main effects and the interaction effects are included. The data sets with only one simulated segment are discarded from the ANOVA analysis here, therefore the ANOVAs are based on 1920 (24 conditions × 10 repetitions × eight methods) observations. Cluster analysis is done with SAS 9.2, the latent class models are estimated with Latent GOLD 4.0.

4.1 Success Rate

In order to get an idea of the success rates for each segmentation method, Table 3 presents the percentages of the estimated numbers of segments and the true simulated numbers. Note that for the segmentation methods based on the *forces*, the *dfbetas* and the *choices*, relatively high success rates are observed in case of one simulated segment. In contrast, the one segment success rates for the *individual-level estimates* with cluster and latent class analysis equal only 35% and 37% respectively. Also in case of two and three segments, the *individual-level partworth estimates* perform poorly as segmentation basis. No success rates higher than 37% are observed. The performance of the *forces* and the *dfbetas* is good for the cluster methods, with success rates between 56% and 72%. For the latent class methods on the other hand, the percentage of correctly estimated numbers of segments drops down to an average of 43%. This is mainly due to the poor results for two and three segments.

The respondents' *choices* seem to perform best with respect to the success

rate. Note that when one segment is simulated respectively 88% and 100% of the data sets were correctly analyzed for the *choices* with cluster and latent class analysis. Also in case of two segments high percentages are observed (68% and 94%) for these methods. Unfortunately, it looks as if the presence of three segments hinders segmentation based on *choices*. Especially for the cluster method the success rate highly decreases (35%). Moreover, note that a lot of three segment data sets were underidentified as two segments (63%). Also for the latent class method a large decrease in success rate is observed (75%).

Nylund et al. (2007) mention the superiority of BIC as a criterion to decide on the number of segments for continuous latent class analysis. In contrast, for the respondents' *choices*, which are discrete indicators, we obtained better segmentation results by the use of the AIC3 information criterion ($AIC3_{LL} = -2LL + 3K$ (Vermunt and Magidson 2005a)). The comparison between the BIC and the AIC3 estimated numbers of segments for the three segment data sets is presented in Table 4. Indeed, an improvement in success rate of 11% is obtained for the *choices* in latent class modeling using the AIC3 criterion. For the *forces*, the *individual-level estimates* and the *dfbetas* on the other hand the use of the AIC3 criterion results in huge decreases in success rate and in extensive overestimation of the number of segments.

The indicators used and the classification method both have a significant main effect and a significant interaction effect on whether or not the true number of segments is uncovered (Figure 1). Besides the mean success rates for the eight segmentation methods, the table in Figure 1 also gives the results of a pairwise comparison of these mean rates. The superscript numbers indicate statistical significant differences (p < 0.01). Suppose the mean success rate of segmentation method *a* is significantly higher than method *b*'s mean rate, a superscript number *b* is attached to the mean success rate of the superior method *a*. The superscript number indicates that method *a* outperforms method *b* with respect to the success rate. The lowest mean success rates are obtained for the *individual-level parameter estimates*. Together with the methods that classify the *forces* and the *dfbetas* into latent classes, these segmentation methods are outperformed by the remaining cluster methods and latent class modeling with *choices*. Dividing the respondents' *choices* into segments with latent class models performs significantly better than any other segmentation method with respect to the success rate. The latter is thus preferred if one needs to decide how many segments there are in a given data set.

The effects on mean success rate of the additional factors included in the simulation setup are given in Table 5. For each factor separately, the mean success rates are compared between factor levels. This for each segmentation method separately and across methods (final column in Table 5). For example, for latent class modeling of the *choices* the mean success rate for all two segment cases (94.2%) is significantly higher (*) than the mean rate for the three segment cases (75.0%). Since the variance factor has three levels, ANOVAs were used to analyze the effect on mean success rate. In case of significant differences between the mean rates for the three variance levels, a superscript (*) is attached to the zero variance level. As expected, fewer segments and a larger mean separation

16

between the segment means of the partworths, increase the success rate of the segmentation methods. A larger within-segment variation on the other hand decreases, though not significantly, the probability of correctly estimating the true number of segments. Overall, there is no significant difference in mean success rate between the small and the large sample size cases. One can conclude that the number of respondents has no general effect on the mean success rate of the segmentation methods. Though, one must be careful drawing conclusions since the discrepancy between the two sample size levels (100 and 300 subjects) could be too small to observe significant effects.

Table 3: The number of estimated segments with respect to the true simulated number of segments (%)

Classification	Indicator	True	1	2	3	4	5	6	7	8	9	10
Cluster	Forces	1	**70**	22	8	0	0	0	0	0	0	0
		2	20	**60**	15	4	1	0	0	0	0	0
		3	6	33	**56**	5	0	0	0	0	0	0
	Individual	1	**35**	53	8	2	2	0	0	0	0	0
		2	49	**26**	17	7	1	0	0	0	0	0
		3	22	25	**32**	14	4	3	0	0	0	0
	Dfbetas	1	**72**	18	8	0	0	2	0	0	0	0
		2	22	**59**	13	5	1	0	0	0	0	0
		3	5	33	**57**	5	0	0	0	0	0	0
	Choices	1	**88**	7	3	0	2	0	0	0	0	0
		2	29	**68**	3	0	0	0	0	0	0	0
		3	2	63	**35**	0	0	0	0	0	0	0
Latent Class	Forces	1	**67**	12	13	0	2	2	2	1	0	1
		2	22	**28**	20	9	6	1	6	4	1	3
		3	7	24	**32**	16	3	6	3	2	2	5
	Individual	1	**37**	25	12	12	3	5	2	2	1	1
		2	14	**29**	15	11	11	8	3	3	4	2
		3	10	24	**20**	14	6	9	5	2	4	6
	Dfbetas	1	**62**	20	8	2	0	5	0	0	2	1
		2	16	**24**	23	7	11	6	7	2	2	2
		3	8	15	**43**	10	4	3	10	5	0	2
	Choices	1	**100**	0	0	0	0	0	0	0	0	0
		2	6	**94**	0	0	0	0	0	0	0	0
		3	0	25	**75**	0	0	0	0	0	0	0

Table 4: The number of estimated segments for the latent class methods with BIC and AIC3 in case of three simulated segments (%)

Classification	Indicator	IC	1	2	3	4	5	6	7	8	9	10
Latent Class	*Forces*	*BIC*	7	24	**32**	16	3	6	3	2	2	5
		AIC3	1	7	**15**	13	7	9	10	11	7	20
	Individual	*BIC*	10	24	**20**	14	6	9	5	2	4	6
		AIC3	0	3	**7**	11	12	6	7	9	17	28
	Dfbetas	*BIC*	8	15	**43**	10	4	3	10	5	0	2
		AIC3	0	7	**17**	9	2	14	14	11	11	15
	Choices	*BIC*	0	25	**75**	0	0	0	0	0	0	0
		AIC3	0	14	**86**	0	0	0	0	0	0	0
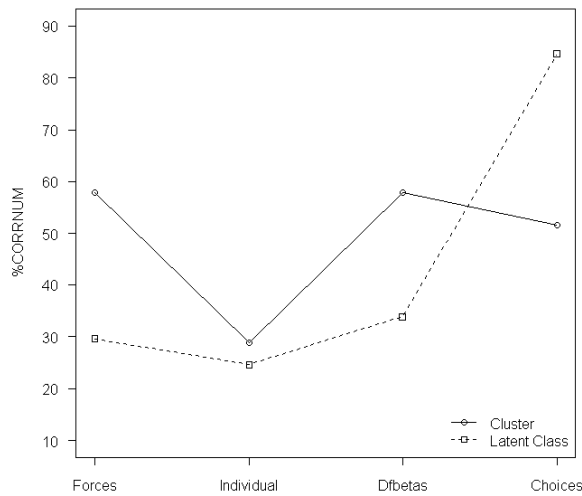
Classification	Indicator		%CORRNUM
Cluster	*Forces*	(1)	$57.9^{2,5,6,7}$
	Individual	(2)	28.8
	Dfbetas	(3)	$57.9^{2,5,6,7}$
	Choices	(4)	$51.7^{2,5,6,7}$
Latent Class	*Forces*	(5)	29.6
	Individual	(6)	24.6
	Dfbetas	(7)	33.8
	Choices	(8)	$84.6^{1,2,3,4,5,6,7}$

Figure 1: Mean success rate for each segmentation method

18

Table 5: Mean success rate under each factor level for each method

Factor		Cluster				Latent Class				Row mean
		Forces	Individual	Dfbetas	Choices	Forces	Individual	Dfbetas	Choices	
Segments	2	60.0	25.8	59.2	68.3*	27.5	29.2	24.2	94.2*	48.5*
	3	55.8	31.7	56.7	35.0	31.7	20.0	43.3*	75.0	43.6
Respondents	100	52.5	25.8	54.2	54.2	34.2	39.2*	46.7*	75.0	47.7
	300	63.3	31.7	61.7	49.2	25.0	10.0	20.8	94.2*	44.5
Variance	0.00	67.5	30.0	66.3	55.0	31.3	23.8	31.3	85.0	48.8
	0.05	53.8	23.8	57.5	56.3	31.3	22.5	41.3	80.0	45.8
	0.10	52.5	32.5	50.0	43.8	26.3	27.5	28.8	88.8	43.8
Separation	1	44.2	35.0*	44.2	35.8	25.0	25.8	26.7	74.2	38.9
	2	71.7*	22.5	71.7*	67.5*	34.2	23.3	40.8*	95.0*	53.3*
		57.9	28.8	57.9	51.7	29.6	24.6	33.8	84.6	

19

4.2 Membership Recovery

As mentioned earlier, for the analysis of membership and parameter recovery the number of clusters or latent classes is set equal to the true simulated number of segments. After classification of the subjects, maximum likelihood estimates of the model partworths are obtained for each cluster or latent class separately. The discrepancies between the true mean partworths of the segments and the estimates of the means in the clusters and latent classes are examined in the following section. First the results for the percentage of correctly classified subjects are described.

For the percentage of correctly classified subjects the main effect of the classification method is not significant on a 0.01 significance level, in contrast to the main effect of the indicators used and the interaction effect between both (Figure 2). The mean percentage of correctly classified subjects for latent class modeling of the *choices* is significantly higher than any other method's mean percentage. Also with respect to membership recovery classifying the respondents' *choices* into segments with latent class models is superior and outperforms the other methods. The second-best methods are the methods that cluster the *forces*, the *dfbetas* and the *choices*. No significant differences in the mean percentages of correctly classified subjects are observed for these cluster methods. They are followed by latent class modeling of the *forces* and the *dfbetas*, which are outperformed by their cluster counterparts and the segmentation methods based on the *choices*. Finally for both cluster analysis and latent class models, the *individual-level partworth estimates* perform worst. Moreover the mean percentage of correctly classified subjects for the *individual-level estimates* cluster method is significantly the lowest across all segmentation methods. Thus not only the success rate but also membership recovery deteriorates for the *individual-level estimates* with both cluster and latent class analysis and for the *forces* and the *dfbetas* in latent class models.

A larger sample size has no significant effect on the percentage of correctly classified subjects. The remaining effects on membership recovery are all as one would expect (Table 6). Fewer segments increase the mean percentage of correctly classified subjects. Membership recovery improves when smaller within-segment variation is considered. Especially in case of no within-segment variation, in comparison to a small amount of variation, significant improvements in mean percentage of correctly classified subjects are observed for the segmentation methods. Finally, a change in the mean separation between the segment means of the partworths largely influences membership recovery. For all segmentation methods significantly more subjects are classified into the correct segment when dissimilar segments, with mean separation 2, are considered instead of similar segments, with mean separation 1. This effect is intuitively clear and again as expected.

20
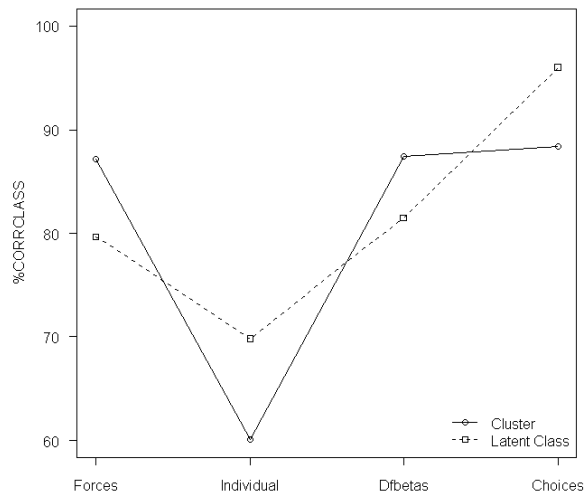
Classification	Indicator		%CORRCLASS
Cluster	Forces	(1)	$87.2^{2,5,6,7}$
	Individual	(2)	60.1
	Dfbetas	(3)	$87.4^{2,5,6,7}$
	Choices	(4)	$88.4^{2,5,6,7}$
Latent Class	Forces	(5)	$79.7^{2,6}$
	Individual	(6)	69.8^{2}
	Dfbetas	(7)	$81.5^{2,6}$
	Choices	(8)	$96.0^{1,2,3,4,5,6,7}$

Figure 2: Mean percentage of correctly classified subjects for each segmentation method

Table 6: Mean percentage of correctly classified subjects under each factor level for each method

| Factor		Cluster				Latent Class				Row mean
		Forces	Individual	Dfbetas	Choices	Forces	Individual	Dfbetas	Choices	
Segments	2	88.3	63.2*	88.7	90.2*	79.9	73.0*	81.8	97.3*	82.8*
	3	86.0	56.9	86.2	86.6	79.5	66.6	81.3	94.8	79.7
Respondents	100	86.7	60.7	86.7	88.5	80.1	69.5	81.9	95.7	81.2
	300	87.7	59.5	88.2	88.2	79.2	70.0	81.1	96.3	81.3
Variance	0.00	90.3*	61.2	89.8	91.2*	81.5	71.9	83.5	97.3	83.3*
	0.05	84.4	58.9	85.0	87.2	78.8	70.0	80.5	95.4	80.0
	0.10	86.9	60.0	87.5	86.7	78.7	67.4	80.6	95.3	80.4
Separation	1	81.8	53.4	82.1	83.1	73.0	61.2	74.3	93.8	75.3
	2	92.6*	66.8*	92.8*	93.7*	86.4*	78.4*	88.7*	98.3*	87.2*
		87.2	60.1	87.4	88.4	79.7	69.8	81.5	96.0	

4.3   Parameter Recovery

For the recovery of the mean partworths of the segments, the ANOVA results are visualized in Figure 3. Note that the RMSE values are an indication of the estimation error when estimating the mean partworths of each segment and thus, in contrast to the success rate and the percentage of correctly classified subjects, smaller values are better. Both the classification method and the segmentation indicators, as well as the interaction between these two factors, significantly affect the recovery of the mean partworths of the segments. The use of the subject *forces* and the *dfbetas* with latent class models performs poorly with respect to coefficient recovery. Both segmentation methods are outperformed by the methods that cluster the *forces*, the *dfbetas* and the *choices* and by latent class modeling of *choices*. For these methods the mean estimation error is almost four times smaller than the mean errors for latent class modeling of the *forces* and the *dfbetas*. Note that the performance of the *individual-level partworth estimates* with respect to parameter recovery is mediocre with latent class models. Its mean RMSE value is lower than the mean value of the *forces* and *dfbetas* but still almost three times the mean value for latent class modeling of *choices*. In contrast, in the cluster cases recovery from the *individual-level estimates* is more than twice as bad in comparison to the remaining indicators.

Mean partworth recovery generally improves for a decrease in the number of segments (Table 7). For the *individual-level estimates* with both cluster and latent class analysis and for the *choices* cluster method, the mean RMSE value for two segments is approximately one third lower than the three segments mean value. For latent class analysis of the *choices* the mean value for three segments is even more than twice the two segment mean. For the majority of the segmentation methods an increase in the number of respondents improves, though not always significantly, parameter recovery. More data give one the ability to estimate model parameters more accurately. Note that an increase in sample size will only correspond to an estimation improvement when a sufficient percentage of subjects is classified into the correct segment and when the misclassified respondents do not influence the estimation results much. Note that for latent class modeling of *choices* augmenting the sample size from 100 to 300 halves the mean RMSE value. The performance of the latter with respect to mean partworth recovery is clearly superior, especially in case of fewer segments and more respondents in the simulated data.

Going from no within-segment variation to a small amount of variation deteriorates parameter recovery, though differences in mean RMSE value between 0.05 and 0.10 variance are negligible. Also for the success rate and the percentage of correctly classified subjects, no significant differences in mean values were observed between the 0.05 and 0.10 level of the variance factor. This indicates that this difference in variance is not large enough to significantly alter the performance of the segmentation methods.

The effects of the number of segments, the number of respondents and the within-component variation on the mean RMSE values are thus as expected. Finally, for all segmentation methods mean parameter recovery is improved for a

smaller mean separation between the partworth means. This result is probably due to the fact that a misclassified subject influences mean partworth estimation more severely in case the dissimilarity between the segments is larger. The true simulated mean parameter values and their estimates will be further apart and the RMSE will increase.
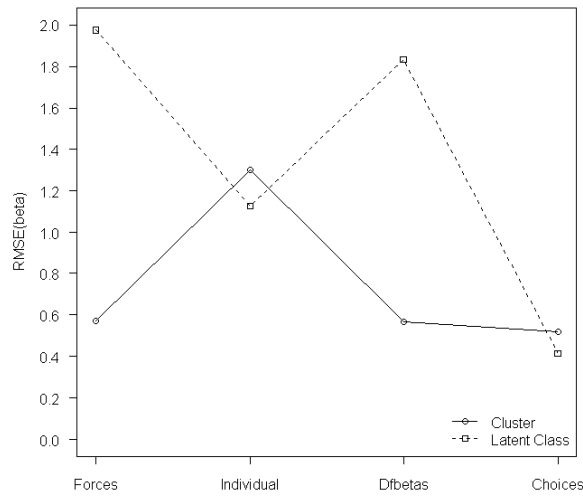
Classification	Indicator		$RMSE(\boldsymbol{\beta})$
Cluster	Forces	(1)	$0.5708^{5,7}$
	Individual	(2)	1.3022
	Dfbetas	(3)	$0.5675^{5,7}$
	Choices	(4)	$0.5194^{5,7}$
Latent Class	Forces	(5)	1.9747
	Individual	(6)	1.1275
	Dfbetas	(7)	1.8332
	Choices	(8)	$0.4145^{5,7}$

Figure 3: Mean estimation error for each segmentation method

Table 7: Mean estimation error under each factor level for each method

Factor		Cluster				Latent Class				Row mean
		Forces	Individual	Dfbetas	Choices	Forces	Individual	Dfbetas	Choices	
Segments	2	0.5014	0.9824*	0.4983	0.4222*	2.1694	0.8677*	1.8971	0.2435*	0.9478
	3	0.6401	1.6220	0.6366	0.6165	1.7800	1.3874	1.7693	0.5856	1.1297
Respondents	100	0.6498	1.6265	0.6516	0.5486	1.6405	0.9854*	1.3267	0.5454	0.9968
	300	0.4918*	0.9779*	0.4834*	0.4901	2.3089	1.2697	2.3397	0.2837*	1.0806
Variance	0.00	0.5886	1.1553	0.5937	0.4438*	1.1791	1.1112	1.3250	0.3986	0.8494
	0.05	0.5649	1.3887	0.5592	0.5478	3.0825	1.1885	2.0888	0.3816	1.2253
	0.10	0.5587	1.3626	0.5494	0.5666	1.6625	1.0828	2.0857	0.4633	1.0415
Separation	1	0.4594*	1.1031*	0.4545*	0.4299*	1.8399	0.9427*	1.2452	0.2800*	0.8443*
	2	0.6821	1.5014	0.6805	0.6089	2.1095	1.3123	2.4212	0.5491	1.2331
		0.5708	1.3022	0.5675	0.5194	1.9747	1.1275	1.8332	0.4145	

The relative performance of the segmentation methods with respect to the three performance measures, visualized in the interaction plots, are similar. This is not a surprise since all three measures represent the segmentation performance of the methods. Only the mean estimation errors for latent class modeling of the *forces* and the *dfbetas* seem to be outlying (Figure 3). In consistency with the other performance measures and corresponding interaction plots the mean parameter recovery is expected to be slightly better for the *forces* and the *dfbetas* than for the *individual-level partworth estimates* with latent class models. After inspection of the data, indeed for some simulated data sets very high estimation errors, which largely increase the mean values, are observed when classifying the *forces* and the *dfbetas* with latent class models. From Figure 2 it is clear that for latent class modeling significantly more subjects are correctly classified with the *forces* and the *dfbetas* than with the *individual-level estimates*. This implies that although relatively many subjects are classified in the correct segment with latent class modeling of the *forces* and the *dfbetas*, the misclassified respondents severely deteriorate mean partworth estimation. This confirms once more the poor performance of latent class segmentation with the *forces* and the *dfbetas* as segmentation basis.

5   Discussion and Conclusions

The aim of the simulation study was to examine the segmentation power of eight two-stage segmentation methods, defined by four types of indicators and two classification methods. Not only their ability to uncover the true number of simulated segments but also the classification performance of the methods, after the number of clusters or classes is set equal to the true simulated number of segments, were compared. The segmentation methods are tested on simulated data sets of respondents' choices for a preset choice design. Although ratings data have been extensively used as segmentation basis in the literature, choice-based conjoint data are used in this research due to the recent popularity of and growing interest in choice experiments. Andrews et al. (2002a) and Karniouchina et al. (2009) point out the lower information content of choice data, with respect to metric conjoint data or preference ratings, therefore it was unclear whether the simulated segments in the data sets could be detected with the proposed methods.

For the *individual-level estimates* of the model partworths low success rates, low classification percentages and high estimation errors are observed. Both the cluster and the latent class method consider the *individual-level estimates* as the subjects' true attribute utilities and ignore any possible estimation error. To the extent that the estimates are accurate, they can serve as a good segmentation basis since homogeneity in consumers' preferences and benefits leads to similar parameter estimates. Unfortunately the potential instability of partworth estimates derived at the individual level, especially with a limited number of choice sets, makes the *individual-level estimates* a poor and unreliable segmentation basis which is confirmed by the simulation results. Note that the

partworths in the conditional logit model were estimated on the individual level with maximum likelihood techniques. Probably more accurate individual-level estimates could be obtained using Hierarchical Bayes techniques for the mixed logit model or the individually sequential Bayesian approach for constructing conjoint choice experiments proposed by Yu (2009). As mentioned before, these more advanced models are very complex and demand more sophisticated software for estimation.

The segmentation performance of the *forces* and the *dfbetas*, both with cluster and latent class classification, are very similar. This similarity is not surprising since both indicators express a subject's influence on the aggregate partworth estimates of the conditional logit model, the former using gradient values of the likelihood function and the latter by means of an outlier detection measure. Their segmentation results are good for the cluster method but deteriorate when used with latent class models. Especially mean partworth recovery is poor for latent class modeling of the *forces* and the *dfbetas*. Note that for the continuous latent class cluster model, class-specific multivariate normal distributions were assumed. Though this assumption seems reasonable for the *individual-level estimates* of the model partworths, it might be too restrictive for the *forces* and the *dfbetas*.

Despite the low information of the *choices*, as only a subject's choice and no information about its strength of preference is given, they seem to perform best with respect to the segmentation performance measures. Particularly with latent class modeling, the superiority of the *choices* is apparent. Nevertheless, some caution is in order when drawing conclusions. First of all, the *choices* mainly perform best in case of ideal experimental conditions such as a few number of segments, a large sample size, small within-component variation and a high mean separation between the partworth means of the segments. Moreover, some calculations show that for the simulation setup at hand the choice probabilities of the alternatives in the choice sets are rather extreme (close to zero or one) for each segment. This dominance of alternatives in segments facilitates segmentation based on the *choices*. Therefore the question rises whether this segmentation basis would be effective for real choice-based conjoint data, for which probably more diversity in choices is observed within each segment.

An additional simulation study confirms these suspicions. A setup with two segments was chosen such that the choice probabilities for the three alternatives in each choice set were more similar for both segments. The interaction plots for the mean success rate (for the latent class methods BIC is used for the *forces*, the *individual-level estimates* and the *dfbetas* whereas AIC3 is used for the *choices*), membership and partworth recovery for the additional study are given in Figure 4.

In the original simulation study there were no significant differences between the mean success rates of the cluster methods with the *forces*, the *dfbetas* and the *choices*. Latent class classification of the *choices* outperformed these cluster methods with respect to mean success rate. The same results were obtained for the mean percentage of correctly classified subjects. But the superiority of latent class modeling of the *choices* is swept away in the new study, its mean

success rate and mean percentage of correctly classified subjects is indifferent to the corresponding means of the cluster methods with the *forces* and the *dfbetas*. As in the original study, the mean success rate and the mean percentage of correctly classified subjects for the *choices* with cluster analysis are significantly lower than the corresponding means of its latent class counterpart.

For mean parameter recovery similar decreases in the segmentation performance of the *choices* are observed. Whereas the mean estimation errors for the *choices* with both cluster and latent class analysis were slightly lower than the mean errors for cluster classification of the *forces* and the *dfbetas* in the original study, mean parameter recovery is somewhat better for the latter in the new study. Nevertheless, for both studies differences in estimation error between these four methods are small and not significant. The additional study confirmed the poor segmentation power of the *individual-level estimates* and latent class modeling of the *forces* and the *dfbetas*.
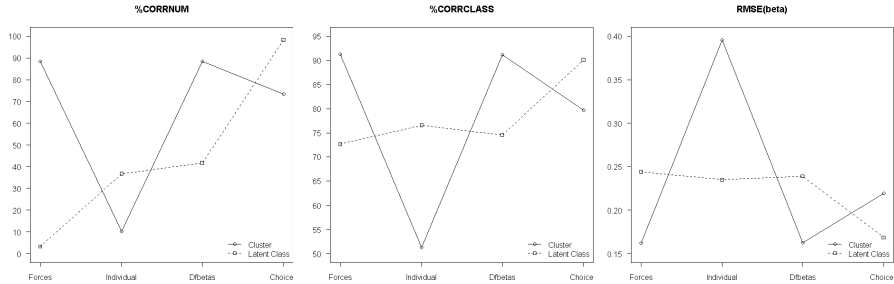

Figure 4: Interaction plots for the additional simulation study

An extensive comparison of the eight two-stage segmentation methods shows that the best results are achieved with cluster classification of the *forces* and the *dfbetas* and with latent class modeling of the *choices*. These segmentation methods perform similarly with respect to success rate, membership and segment mean partworth recovery. Their segmentation results are encouraging and reveal the segmentation ability of more straightforward two-stage segmentation methods. Nevertheless, a main drawback of the cluster segmentation methods is the lack of good criteria to decide on the number of clusters. This decision consequently is a manual, based on plots of some cluster evaluation statistics, and therefore subjective task. Because of this inevitable, and at this point unresolved, problem, cluster success rates depend on the researcher evaluating the plots. Additionally cluster results depend highly on the cluster method used (Desarbo et al. 1992). Due to these drawbacks of cluster analysis and the promising results for latent class classification of the *choices*, one could argue that there is no need to compute the *forces* and the *dfbetas* and that one can rely solely on the respondents' *choices* in order to find consumer segments. On the other hand, the methods based on the *choices* are favored because the *choices*' dimension (16) is a lot higher than the dimension of the *forces* and

the *dfbetas* (6). But one might not always be able to obtain a large enough number of choices for each respondent separately. Despite these shortcomings, there is no reason to turn down the use of two-stage segmentation methods to uncover consumer segments. Though we recommend future research by means of additional simulation studies and applications on real-life choice experiments, this research supports the segmentation ability of the two-stage segmentation methods clustering respondents' *forces* and *dfbetas* and classifying their *choices* with latent class analysis.

Acknowledgements

Appendix A

Table 8: Design of the simulation study

Choice set	Alt	Attr A	B	C		Choice set	Alt	Attr A	B	C
1	I	2	1	3		9	I	1	2	2
	II	3	2	1			II	3	1	1
	III	1	3	2			III	2	3	3
2	I	3	3	3		10	I	3	2	3
	II	1	1	1			II	2	3	2
	III	2	2	2			III	1	1	1
3	I	3	1	2		11	I	3	1	2
	II	2	3	1			II	2	2	1
	III	1	2	3			III	1	3	3
4	I	2	3	2		12	I	2	1	3
	II	1	1	3			II	3	3	1
	III	3	2	1			III	1	2	2
5	I	2	1	1		13	I	2	3	3
	II	3	3	3			II	3	1	2
	III	1	2	2			III	1	2	1
6	I	3	1	2		14	I	1	3	2
	II	1	3	1			II	2	1	1
	III	2	2	3			III	3	2	3
7	I	1	3	1		15	I	1	1	3
	II	3	2	3			II	2	2	2
	III	2	1	2			III	3	3	1
8	I	1	1	3		16	I	3	3	1
	II	2	2	1			II	1	2	3
	III	3	3	2			III	2	1	2

Appendix B

Table 9: Scatterplots of the scores for the first two principal components (number of segments (Seg), number of respondents ($Resp$), variance (Var) and mean separation between segment means (Sep)). The first principal component is on the horizontal axis.

2 Seg				
		$Resp$	100	300
Sep	Var			
1	0.05			
	0.10			
2	0.05			
	0.10			

31

Table 10: Scatterplots of the scores for the first two principal components (number of segments (*Seg*), number of respondents (*Resp*), variance (*Var*) and mean separation between segment means (*Sep*)). The first principal component is on the horizontal axis.

3 *Seg*			
	Resp	100	300
Sep	*Var*		

		100	300
1	0.05		
	0.10		
2	0.05		
	0.10		

32

References

[1] Allenby, G.M., & Rossi, P.E. (1999). Marketing Models of Consumer Heterogeneity. *Journal of Econometrics*, 89, 57-78.

[2] Allenby, G.M., Rossi, P.E., & McCulloch, R.E. (2005). Hierachical Bayes Models: a Practitioners Guide. Ssrn working paper, Ohio State University, University of Chicago.

[3] Alriksson, S., & Öberg, T. (2008). Conjoint Analysis for Environmental Evaluation - A Review of Methods and Applications. *Env Sci Pollut Res*, 15, 244-257.

[4] Andrews, R.L., Ainslie, A., & Currim, I.S. (2002a). An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity. *Journal of Marketing Research*, 39, 479-487.

[5] Andrews, R.L., Ansari, A., & Currim I.S. (2002b). Hierarchical Bayes Versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction and Partworth Recovery. *Journal of Marketing Research*, 39, 87-98.

[6] Andrews, R.L., & Currim, I.S. (2003). A Comparison of Segment Retention Criteria for Finite Mixture Logit Models. *Journal of Marketing Research*, 40, 235-243.

[7] Biggadike, E.R. (1981). The Contributions of Marketing to Strategic Management. *The Academy of Management Review*, 6, 621-632.

[8] Bigsby, H., & Ozanne, L.K. (2002). The Purchase Decision: Consumers and Environmentally Certified Wood Products. *Forest Products Journal*, 52, 100-105.

[9] Cattin, P., & Wittink, D.R. (1982). Commercial Use of Conjoint Analysis: A Survey. *Journal of Marketing*, 46, 44-53.

[10] Chaim, M., Soutar, G., & Yeo, A. (2009). Online and Off-line Travel Packages Preferences: A Conjoint Analysis. *International Journal of Tourism Research*, 11, 31-40.

[11] Chrzan, K., & Orme, B. (2000). An Overview and Comparison of Design Strategies for Choice-Based Conjoint Analysis. *Sawtooth Software Research Paper Series*.

[12] Colombo, S., Hanley, N., & Louviere, J. (2008). Modeling Preference Heterogeneity in Stated Choice Data: An Analysis for Public Goods Generated by Agriculture. *Stirling Economics Discussion Paper*, 2008-2028.

[13] Cunningham, C.E., Deal, K., Rimas, H., Chen, Y., Buchanan, D.H., & Sdao-Jarvie, K. (2009). Providing Information to Parents of Children with Mental Health Problems: A Discrete Choice Conjoint Analysis of Professional Preferences. *J Abnorm Child Psychol*, 37, 1089-1102.

[14] Currim, I.S. (1981). Using Segmentation Approaches for Better Prediction and Understanding from Consumer Mode Choice Models. *Journal of Marketing Research*, 18, 301-309.

[15] Desarbo, W.S., Wedel, M., Vriens, M., & Ramaswamy, V. (1992). Latent Class Metric Conjoint Analysis. *Marketing Letters*, 3, 273-288.

[16] Desarbo, W.S., Ramaswamy, V., & Cohen, S.H. (1995). Market Segmentation with Choice-Based Conjoint Analysis. *Marketing Letters*, 6, 137-147.

[17] Dolnicar, S. (2002). A Review of Unquestioned Standards in Using Cluster Analysis for Data-Driven Market Segmentation. CD Conference Proceedings of the Australian and New Zeeland Marketing Academy Conference 2002.

[18] Elrod, T., Louviere, J.J., & Davey, K.S. (1992). An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models. *Journal of Marketing Research*, 29, 368-377.

[19] Everitt, B.S., Landau, S., & Leese, M. (2001). *Cluster Analysis*. Arnold, London.

[20] Goodman, L.A. (1974). The Analysis of Systems of Qualitative Variables when some of the Variables are Unobservable: Part I - A Modified Latent Structure Approach. *American Journal of Sociology*, 79, 1179-1259.

[21] Green, P.E., & Krieger, A.M. (1991). Segmenting Markets with Conjoint Analysis. *Journal of Marketing*, 55, 20-31.

[22] Green, P.E., & Srinivasan, V. (1978). Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research*, 5, 103-123.

[23] Green, P.E., & Srinivasan, V. (1990). Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. *Journal of Marketing*, 54, 3-19.

[24] Greene, W.H., & Hensher, D.A. (2003). A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. *Transportation Research Part B*, 37, 681-698.

[25] Gupta, S., & Chintagunta, P.K. (1994). On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models. *Journal of Marketing Research*, 31, 128-136.

[26] Hagenaars, J.A., & McCutcheon, A.L. (eds.) (2002). *Applied Latent Class Analysis*. Cambridge: Cambridge University Press.

[27] Hagerty, M.R. (1985). Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis. *Journal of Marketing Research*, 22, 168-184.

[28] Haley, R.I. (1968). Benefit Segmentation: A Decision-Oriented Research Tool. *Journal of Marketing*, 32, 30-35.

[29] Johnson, R.M. (1971). Market Segmentation: A Strategic Management Tool. *Journal of Marketing Research*, 8, 13-18.

[30] Kamakura, W.A. (1988). A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments. *Journal of Marketing Research*, 25, 157-167.

[31] Karniouchina, E.V., Moore, W.L., van der Rhee, B., & Verma, R. (2009). Issues in the Use of Ratings-Based versus Choice-Based Conjoint Analysis in Operations Management Research. *European Journal of Operational Research*, 197, 340-348.

[32] Kessels, R., Goos, P., & Vandebroek, M. (2006). A Comparison of Criteria to Design Efficient Choice Experiments. *Journal of Marketing Research*, 43, 409-419.

[33] Lancsar, E., & Louviere, J.J. (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. *Pharmacoeconomics*, 26, 661-677.

[34] Lattin, J.M., Carroll, J.D., & Green, P.E. (2003). *Analyzing Multivariate Data*. Thomson Learning Inc.

[35] Lenk, P.J., & Desarbo, W.S. (2000). Bayesian Inference for Finite Mixtures of Generalized Linear Models with Random Effects. *Psychometrika*, 65, 93-119.

[36] Louviere, J.J., & Woodworth, G. (1983). Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. *Journal of Marketing Research*, 20, 350-367.

[37] Magidson, J., & Vermunt, J.K. (2002). Latent Class Modeling as a Probabilistic Extension of K-means Clustering. *Quirks Marketing Research Review*, 20, 77-80.

[38] Magidson, J., & Vermunt, J.K. (2004). Latent class analysis, D. Kaplan (ed.), *The sage handbook of quantitative methodology for the social sciences*. Chapter 10 175-198. Thousand Oakes: sage publications.

[39] Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling*, 14, 535-569.

[40] Orme, B., & Howell, J. (2009). Application of Covariates within Sawtooth Software's CBC/HB Program: Theory and Practical Example. *Sawtooth Software Research Paper Series*

[41] Punj, G., & Stewart, D.W. (1983). Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*, 20, 134-148.

[42] SAS Institute Inc. (2009). JMP(r) 8 Statistics and Graphics Guide, 2nd Ed. Cary, NC: SAS Institute Inc.

[43] Sawtooth Software Inc. (2008). The CBC Technical Paper (version 6). *Sawtooth Software Technical Paper Series* 1-27. Retrieved from http://www.sawtoothsoftware.com/download/techpap/cbctech.pdf.

[44] Sharma, S. (1996). *Applied Multivariate Statistics*. John Wiley & Sons, Inc.

[45] Smith, W.R. (1956). Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of Marketing*, 21, 3-8.

[46] Song, J., Jang, T., & Sohn, S.Y. (2009). Conjoint Analysis for IPTV Service. *Expert Systems with Applications*, 36, 7860-7864.

[47] Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

[48] Vermunt, J.K., & Magidson, J. (2002). Latent class cluster analysis. J.A. Hagenaars and A.L. McCutcheon (eds.) *Applied latent class analysis* 89-106. Cambridge: Cambridge university press.

[49] Vermunt, J.K., & Magidson, J. (2005a). Technical Guide for Latent GOLD 4.0: Basic and Advanced. Belmont Massachusetts: Statistical Innovations Inc.

[50] Vermunt, J.K., & Magidson, J. (2005b). Technical Guide for Latent GOLD Choice 4.0: Basic and Advanced. Belmont Massachusetts: Statistical innovations Inc.

[51] Vriens, M., Wedel, M., & Wilms, T. (1996). Metric Conjoint Segmentation Methods: A Monte Carlo Comparison. *Journal of Marketing Research*, 33, 73-85.

[52] Wedel, M., & Desarbo, W.S. (1995). A Mixture Likelihood Approach for Generalized Linear Models. *Journal of Classification*, 12, 21-55.

[53] Wind, Y. (1978). Issues and Advances in Segmentation Research. *Journal of Marketing Research*, 15, 317-337.

[54] Wittink, D.R., & Cattin, P. (1989). Commercial Use of Conjoint Analysis: An Update. *Journal of Marketing*, 53, 91-96.

[55] Wittink, D.R., Vriens, M., & Burhenne, W. (1994). Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections. *International Journal of Research in Marketing*, 11, 42-52.

[56] Wolfe, J.H. (1970). Pattern Clustering by Multivariate Mixture Analysis. *Multivariate Behavioral Research*, 5, 329-350.

[57] Yu, J., Goos, P., & Vandebroek, M. (2009). Individually Adapted Sequential Bayesian Designs for Conjoint Choice Experiments. Available at SSRN: http://ssrn.com/abstract=1395663.