Finite Mixture Models in Market Segmentation: A Review and Suggestions for Best Practices

Michael Tuma and Reinhold Decker
Bielefeld University, Bielefeld, Germany
mtuma@uni-bielefeld.de
rdecker@wiwi.uni-bielefeld.de

Abstract: Recently, Andrews, Brusco and Currim (2010) noted that some of the hesitancy on the part of practitioners to adopt model-based (MB) methods in market segmentation (MS) may stem from an insufficient awareness of their performance relative to their non-model-based (NMB) counterparts. Comparisons of MB and NMB methods should provide business researchers with information as to precise conditions in which the former should be preferred. Moreover, finite mixture models (FMMs) have grown in their use since 2000 and, as there is no recent survey-based empirical literature examining their application, a comprehensive review of their usage in segmentation research seems to be of use. This article discusses some of the critical issues involved when using FMMs to segment markets, takes a closer look at comparison simulation studies in order to highlight conditions under which a business analyst might consider the application of an FMM approach, discusses model selection as well as validation issues and provides suggestions for best practices and potential improvements. Furthermore, it presents an empirical survey that seeks to provide an up-to-date assessment of FMM application in MS.

Keywords: market segmentation, model-based clustering, finite mixture models, latent class models

1. Introduction

Marketers usually address consumer heterogeneity by grouping consumers into segments consisting of those consumers having relatively similar product or service needs. Cluster analysis (CA) is one of the most widely used methods in segmenting markets (Wedel and Kamakura, 2000). Most clustering done in MS practice is based largely on heuristic procedures like Ward's method and *k*-means (Tuma, Decker, and Scholz, 2011). However, the often insufficient statistical basis of such methods appears to be a major drawback for their use and crucial issues in segmentation, such as determining the optimal number of segments, can hardly be answered by heuristic procedures. Lubke and Muthén (2005:23), for instance, note in this respect that "clustering using *k*-means is achieved based on an arbitrarily chosen criterion which aims at minimizing the within-cluster variability while maximizing between-cluster variability".

FMMs or MB clustering are a principal alternative to heuristic-based algorithms. They are "viewed as elegant procedures that incorporate mixtures of parametric distributions to define the true cluster structure" (Steinley and Brusco, 2011:63). They are the main statistical approach to clustering and segmentation and some academic literature tout and advocate their usage "as a preferred approach because of the provision of a formal statistical model" (Andrews, Brusco and Currim 2010:609; McLachlan and Peel, 2000). In practice, a business analyst embarking on the use of FMMs for segmentation has to deal with several crucial issues like selecting the type of FMM, variable selection, determination of the number of variables and sample size, data pre-processing requirements, determination of the number of segments, validity and stability tests of the obtained results as well as the interpretation and substantial description of segments. All these issues highly influence the outcome and quality of the derived market segments regarding further market-centric activities. The seminal work of Wedel and Kamakura (2000) provides a comprehensive review of FMM applications in MS until the turn of the last century. Since then, however, there have been many new developments in FMMs applications. In the domain of marketing, new models have been designed, implemented and published in top-tier journals (e.g. Hahn et al., 2002). Andrews, Brusco and Currim (2010) suggest that some of the hesitancy on the part of practitioners to apply MB approaches to MS may stem from an insufficient knowledge of their performance relative to their NMB counterparts. Whereas the performance of NMB techniques has been evaluated using Monte Carlo simulations for more than three decades (Blashfield, 1976; Steinley and Brusco, 2008a/2008b), extensive simulation studies comparing MB and NMB methods are still slowly emerging in the literature and should ultimately provide business researchers with information as to precise conditions under which MB approaches are apt to be preferred. Moreover, as user friendly software packages like Latent GOLD have made their debut and FMMs have grown in their use since 2000 and, to the best of our knowledge, as

Reference this paper as: Tuma, M and Decker, R."Finite Mixture Models in Market Segmentation: A Review and Suggestions for Best Practices" *The Electronic Journal of Business Research Methods* Volume 11 Issue 1 2013 (pp 02-15), available online at www.ejbrm.com

there is no recent survey-based empirical literature examining their managerial application, an up-to-date review of their usage in segmentation research seems to be overdue.

Against this backdrop, the aim of this article is two-fold: First, it reviews the crucial issues in MS using FMMs. In order to inform business managers and marketing researchers about the performance of competing methods, it builds on the suggestion of Andrews, Brusco and Currim (2010) and, drawing on prior findings in marketing and other disciplines, provides a closer look at comparison simulation studies, with set-ups and key results regarding MB and NMB segmentation methods. The paper contributes to the current literature by integrating previous research results in order to present a holistic overview and comprehensive insights (e.g. concerning the performance of various types of FMMs) into new developments by addressing the critical issues. Furthermore, it presents an empirical study that investigates how some of the important problems in MS using FMMs have been addressed by researchers and, last but not least, it examines how the application of easy-to-use software packages have led to more rigorous applications of FMMs in MS.

Accordingly, the remainder of the paper is structured as follows. Section 2, among other things, takes a closer look at some of the critical issues when applying FMMs. Section 3 presents selected simulation studies. Section 4 specifies the research questions and the methodology underlying our study, whereas Section 5 is devoted to important results. This is followed in Section 6 by a brief conclusion, managerial implications and an outlook on future research.

2. Finite mixture models

Several extensions of the basic FMM approach have been suggested in the recent past. Finite mixture regression models (FMRMs) (Wedel and Kamakura, 2000) for example, are able to simultaneously derive segments and segment-specific weights that relate an outcome or dependent variable (e.g. product recommendation or rating) to a set of independent or explanatory variables (e.g. price of a product and product quality) and derive a unique regression model for each segment. Finite mixture approaches have also been developed that combine the strengths of the partial least squares (PLS) method or the covariance structure analysis that are used for better understanding heterogeneity within structural equation models with the advantages of classifying market segments according to FMMs. Jedidi, Jagpal and DeSarbo (1997) pioneered the development of the finite mixture structural equation model (FIMIX-SEM), an approach that combines FMMs, the Expectation-Maximization (EM) algorithm and covariance-based SEM. The original technique is inappropriate for PLS analysis because of divergent methodological assumptions. For this reason, Hahn et al. (2002) introduced the finite mixture partial least squares (FIMIX-PLS) method that combines a FMM procedure with an EM algorithm specifically coping with the ordinary least squares (OLS)-based predictions of PLS. Conceptually, FIMIX-PLS is equivalent to a mixture regression approach. However, the main difference is that the structural model can comprise a multitude of (interrelated) endogenous latent variables (Hahn et al., 2002). One of the assumptions of the aforementioned FMMs is that the available sample has only a single level, i.e., it consists of a sample of independent units, an assumption that is inadequate when the sample to be analyzed has multiple levels, i.e., when units are nested within clusters sharing common environments, experiences and interactions (Lukočiene, Varriale and Vermunt, 2010). The multilevel latent class (MLC) model with finite mixture distributions at multiple levels of a hierarchical structure has been developed for the analysis of data sets having such a multilevel structure.

2.1 Estimation of FMMs

Usually, the FMM parameters are unknown and have to be estimated from the data. There is a remarkable variety of estimation methods such as the method of moments, maximum likelihood (ML), minimum chi-square, and Bayesian approaches (McLachlan and Peel, 2000). Numerical methods for obtaining ML estimates primarily involve the use of gradient methods like Newton-Raphson, quasi-Newton, and Fisher's scoring. Other approaches rely on the EM algorithm (Dempster, Laid and Rubin, 1977), stochastic EM (SEM) or Markov Chain Monte Carlo (MCMC) methods. Parameter estimation in Bayesian methods is with MCMC using Gibbs sampler and Metropolis-Hastings algorithm.

The primary advantages of numerical optimization procedures are their speed, relative to the EM algorithm, and their ability to obtain standard errors for parameter estimates. The primary advantages of the EM algorithm are that (1) at each stage of the iterative process the likelihood is monotonically increasing and (2) under certain regularity conditions, the sequence of likelihoods will converge to at least a local maximum. The latter is an iterative, hill-climbing procedure whose performance can depend severely on the particular starting point (McLachlan and Peel, 2000). Hence, numerous initialization procedures have been suggested in the literature (see, e.g., Melnykov and Maitra, 2010).

Distributional assumptions for the variables have to be made by the researcher when estimating model parameters. Checking the empirical distributions and consulting skewness and kurtosis measures or plotting the data for a visual representation can be very helpful in this regard (Wedel and DeSarbo, 2002).

Identifiability is an issue related to parameter estimation and determines whether a unique solution can be obtained. It can be investigated by inspecting the Hessian matrix of second derivatives of the likelihood (Bekker, Merckens and Wansbeek, 1994). Positive eigenvalues of the information matrix provide evidence of identifiability (Wedel and DeSarbo 2002). The non-identifiability of the components leads to so-called label switching. If this occurs, summary statistics of the marginal distributions will give inaccurate estimates (Dias and Wedel, 2004). Label switching can be detected through investigating iteration plots of the MCMC sampler (Ebbes, Grewal and DeSarbo, 2010). Frühwirth-Schnatter (2006) provides an overview of approaches to address label switching.

2.2 Model selection

2.2.1 Determining the Number of Segments

When applying FMMs to empirical data, the actual number of segments S is unknown and must be inferred from the data itself. The majority of methods devoted to estimating S can broadly be divided into two categories. The first group of methods relies on testing procedures while the second one is based on information criteria (IC). Particularly, the latter class of methods is frequently used for investigating the number of clusters (Sarstedt 2008). These methods determine the number of segments by minimizing the negative log-likelihood function augmented by some penalty function, which increases with the number of parameters and/or the number of observations to reflect its complexity. Table 1 presents some of the model selection criteria that have recently been used in MS.

Table 1: Some model selection criteria[1]

Criterion	Reference
Akaike IC (AIC)	Akaike (1974)
Bayesian IC (BIC)	Schwarz (1978)
Consistent AIC (CAIC)	Bozdogan (1987)
AIC3	Bozdogan (1994)
Normalized Entropy Criterion (NEC)	Celeux and Soromenho (1996)
Validation sample log-likelihood (LOVLG)	Andrews and Currim (2003a)
Sample size adjusted BIC (ssBIC)	Sclove (1987)
Classification error	Garver, Williams and Taylor (2008)
Markov switching criterion (MSC)	Ebbes, Grewal and DeSarbo (2010)
R^2	Garver, Williams and Taylor(2008)
Log marginal density (LMD)	Hofstede, Wedel and Steenkamp (2002)
Deviance IC (DIC)	Spiegelhalter et al. (2002)

2.2.2 Variable selection

Selecting the appropriate clustering variables actually used in segmentation is one of the most fundamental steps in the segmentation process. It has long been recognised that not all variables contribute equally to defining the underlying segment structure. In many multivariate datasets, for example, some of the variables are highly correlated with the others or just do not carry much additional information about the potential segments. Since the performance of segmentation algorithms can be severely affected by the presence of such variables, their elimination can potentially improve both estimation and clustering performance (Melnykov and Maitra, 2010).

[1] Most of the model selection simulation studies focus on IC. Sarstedt et al. (2011) consider classification criteria such as complete log-likelihood, Entropy criterion, etc.

The marketing researcher must also decide on the number of variables to be considered, the sample size and the relation between these and the resulting segments. The relationship between the number of objects to be grouped and the number of variables to be used is important, given that the number of variables used determines the dimensionality of the space within which the method or model is searching for groupings. Formann (1984) suggested a sample size of 2^k, where k represents the number of variables used in segmentation as a rough guide for this relation. Preferably, it should be $5 \cdot 2^k$ respondents.

Special care must be taken with LC regression (LCR) to ensure that the appropriate sample size exists for each segment in the model. Consistent with multiple regression, LCR generally requires at least five observations per independent variable per segment (Hair et al. 1995). A rule of thumb suggests that a minimum sample size of 30 observations/respondents per segment may be adequate (Garver, Williams and Taylor, 2008).

2.3 Other Important Issues

2.3.1 Stability and Validity

Various approaches, subsumed under the terms validity and stability, have been developed to analyse the quality of the final segment solution. The latter can be evaluated by running several clustering procedures or one clustering procedure several times (with different specifications) on the same dataset and testing whether the partitions remain constant and are thus stable. Validation includes attempts by the researcher to ensure that the segments are representative of the general population. Strategies for validation may be based on external, internal and relative criteria (Wedel and Kamakura, 2000).

2.3.2 Software packages

The R environment provides several packages for estimating mixture models, e.g. MCLUST for mixtures of multivariate Gaussian distributions, fpc for mixtures of linear regression models, mmlcr for mixed-mode latent class analysis, poLCA for polytomous outcome variables, and flexmix for FMRMs. In SAS, the package PROC NLP can be used to specify different mixture models. The Stata package fmm estimates FMRMs.

3. Simulation studies

3.1 FMMs

Andrews, Currim and Leeflang (2010) present a simulation experiment that provides insights into the conditions under which prediction bias may occur, and, when it occurs, to understand why by determining the effects of data aggregation (panel vs. store level), heterogeneity, endogeneity, and the number of households, stores, and weeks on bias in sales response predictions. Among other things, they manipulated the degree of heterogeneity between and within segments as well as the number of weeks, stores and households per store considered. Using choice data comprising more than 300 panel and store-level datasets and assuming two segments of consumers for all datasets, they estimated several nested logit models and nested logit FMMs. In this study models explaining within-store heterogeneity (i.e., heterogeneity across store visits) using random distributions for the coefficients produced predictions that were significantly more accurate than those of the other models. One implication of this finding is that if the sole objective of an analysis is to predict segment-specific market response to a new promotional environment, store-level data should suffice. Often, the latter is cheaper to obtain, more widely available, and more computationally efficient than panel data.

Andrews et al. (2010) compare the effectiveness of statistical MB clustering methods with that of more commonly used NMB procedures. They therefore manipulated the number of segments, consumers and characteristic variables, as well as the concordance between response-based and characteristic-based segments, the scale type for characteristics and the availability of predictor variables. From more than 800 generated datasets they found that if a manager's primary objective is to forecast responses for segments of holdout consumers for whom only characteristics are available, NMB procedures perform better than MB procedures. However, if it is important to understand the true segmentation structure in a market as well as the nature of the regression relationships within segments, the MB procedure is clearly preferred.
Andrews, Brusco, and Currim (2010) compare three approaches for forming a consensus segmentation scheme, namely clique partitioning, the SEGWAY algorithm and a method based on a latent class model.

Among others, they manipulated the number of customers and latent segments, the segment membership probabilities as well as the number of partitions and the number of classes within each partition. Using ANOVA to investigate the results of a total of 648 generated datasets, they found that the FMM approached yielded better average recovery of holdout validation segments than did the deterministic methods. For marketing researchers seeking to obtain consensus segmentation, FMM seems to be a promising option.

Andrews, Ainslie and Currim (2002) compare the performance of FM logit models and hierarchical Bayes-estimated mixed logit models with discrete versus continuous representations of heterogeneity in terms of the accuracy of household-level parameters, fit, and forecasting accuracy. The authors experimentally manipulated the number of mixture components, the separation between mixture components, the distribution and variance of coefficients within components, the number of households and purchases as well as the error variance. The set of predictors included one continuous variable (price) and two binary variables (store feature advertisement and aisle display). Based on 288 choice datasets the FMMs proved to have the best overall performance with regard to parameter recovery. In general, the models fit better when the separation between components is larger, the within-component distribution is normal, the within components variance is smaller, and when there is less error variance. The results indicate that FMM is the preferable method to use for marketing analysts seeking to obtain forecasting accuracy.

3.2 Estimation of FMMs

Using a synthetic dataset Dias and Wedel (2004) provide simulation comparisons of ML estimation methods (EM, SEM, MCMC) on the basis of three convergence criteria and initialization methods such as starting with random centers (RC) based on McLachlan and Peel (2000:55), or with a random partition (RP) of the data. Furthermore, they investigated an approach to minimize the label-switching effect based on imposing identifiability constraints on the parameters (e.g. segment sizes, means and variances) and the methods proposed by Celeux (1998), Celeux, Hurn and Robert (2000) (CHR) and Stephens (1997).

For the EM algorithm, the relative criterion and the Aitken's absolute criterion underperformed the absolute criterion. They found that using RC for each convergence criterion decreases the proportion of false solutions (for a given number of iterations). For SEM, RP solutions outperformed the best EM solution in terms of the log-likelihood value. SEM proves to be faster and displays better convergence properties, but is less stable than the EM. For MCMC, the identifiability constraints negatively affect parameter recovery, in particular when the component sizes or variances are constrained. Furthermore, the absence of any label-switching procedure outperformed procedures in which identifiability constraints are imposed and Stephens (1997) and CHR relabeling procedures were the most effective.

Finally, Dias and Wedel (2004) conclude that MCMC is preferable over EM and SEM in recovering the parameters of mixture models. The absolute convergence criterion should be used in conjunction with the EM algorithm and RP with the SEM. Identifiability constraints should be avoided, but, if they have to be used, then the better performing ones, for instance that of Stephens (1997) and CHR, are recommended.

3.3 Model selection

3.3.1 Determining the number of segments

Simulation studies in the context of FMMs can be broadly classified according to the type of FMMs used in generating the datasets. Tables 2, 3 and 4 present set-ups and key results of model selection simulation studies in the contexts of FIMIX-SEM and -PLS, LC models and FMRMs, respectively.[2]

Table 2: Set-ups and key results for model selection simulation studies in the context of FIMIX-SEM and -PLS

	Sarstedt et al. (2011)	Henson, Reise and Kim (2007)	Jedidi, Jagpal and DeSarbo (1997)
Model	PLS	SEM	SEM
No. of manipulated data characteristics	6	5	2
No. of segments	2 or 4	1, 2 or 3	2 or 4

[2] Because of space limitations, we present only the best criteria. The interested reader should consult the respective studies for a complete list of the criteria considered.

	Sarstedt et al. (2011)	Henson, Reise and Kim (2007)	Jedidi, Jagpal and DeSarbo (1997)
Model	PLS	SEM	SEM
Disturbance term of the endogenous latent variables	10% or 25%	-	-
Distance between segment-specific path coefficients (separation between segments)	0.25 or 0.75	-	-
Sample size	100 or 400	500, 1500 or 2500	-
Model complexity	low or high	-	-
Relative segment sizes	Balanced or unbalanced	unbalanced	-
Factor level combinations	64	-	-
Mixture proportion	-	50%-50%, 70%-30%, 90%-10%	-
No. of indicators	-	-	3 or 6
No. of datasets	-	121500	-
Type of dataset	-	-	structured data
Distribution for datasets	-	binomial distribution	multivariate normal
Best criterion	AIC3, CAIC	ssBIC	BIC

Table 3: Set-ups and key results for model selection simulation studies in the context of LC models

	Dias (2006)	Lukočiene and Vermunt (2010)	Lukočiene, Varriale and Vermunt (2010)	Lukočiene, Varriale and Vermunt (2010)
Model	Binary LC	MLC	MLC with categorical indicators	MLC with continuous indicators
No. of manipulated data characteristics	5	3	7	6
No. of variables	5 or 8	6	6 or 10	6
No. of segments	2 or 3	3 at individual level, 2 or 3 at higher level	2 or 3 at both levels	2 or 3 at both levels
Segment sizes	equal proportion, unbalanced	-	0.7 or 0.8	0.7 or 0.8
Separation between segments	well, moderately or weakly separated	Lower level (moderately separated), higher level (from very low to very high)	from very low to very high separation	from very low to very high separation
Sample sizes	600, 1200 or 2400	Lower level (5, 10, 15, 20), higher level (50 or 500)	lower level (5, 10, 20, 50), higher level (30, 100, 1000)	lower level (5, 10, 20, 50), higher level (30, 100, 1000)
Measurement level of variables	-	discrete	discrete	-
No. of datasets	10800	2000	2880	-
Type of dataset	binary	binary with hierarchical or multilevel structure	hierarchical or multilevel structure	hierarchical or multilevel structure
Best criterion	AIC3, AIC	AIC3	AIC3	AIC3, BIC(K)

Table 4: Set-ups and key results for model selection simulation studies in the context of logit and FMRMs

	Andrews and Currim (2003a)	Andrews and Currim (2003b)	Sarstedt and Schwaiger (2007)	Sarstedt (2008)
No. of manipulated data characteristics	7	8	-	5
No. of segments	2 or 3	2 or 3	3	2

	Andrews and Currim (2003a)	Andrews and Currim (2003b)	Sarstedt and Schwaiger (2007)	Sarstedt (2008)
No. of manipulated data characteristics	7	8	-	5
No. of segments	2 or 3	2 or 3	3	2
Regression coefficients in each segment	-	-	S1 (1, 1, 1.5, 2.5); S2 (1, 2.5, 1.5, 4); S3 (2, 4.5, 2.5, 4)	S1 (1, 1,1.5, 2.5); S2 (1, 2.5, 1.5, 4)
Mean separation between segments	small (1.0), medium (1.5) or large (2.0)	small (0.5), medium (1.0) or large (1.5)	-	-
Sample size	100 or 300	100 or 300	Varied in 100-step intervals of [100:1000]	Varied in 10-step intervals of [50:500]
Mean no. of purchases per household	5 or 10	-	-	-
No. of observations per individual	-	5 or 10	-	-
No. of choice alternatives	3 or 6	-	-	-
No. of predictors	-	3 or 6	3	-
Error variance	1.645 or 50% higher	20% (R^2 = 0.80) or 60% (R^2 = 0.40)	-	-
Segment size	5%-10%, 10%-20% or 20% -30%	5%-10%, 10% - 20% or 20%-30%	(0.1, 0.1, 0.8), (0.2, 0.2, 0.6) or (0.3, 0.3, 0.4)	(0.1, 0.9), (0.2, 0.8), (0.3, 0.7), (0.4, 0.6) or (0.5, 0.5)
Measurement level of predictors	-	continuous or discrete	continuous	-
No of datasets	864	1728	-	230000
Type of dataset	scanner panel data	normal data	normal data	normal data
Distribution for datasets	gamma distribution	normal distribution	-	standard normal
Best criterion	AIC3	AIC3	AIC3	AIC3

These tables show that AIC3 performs well. Sarstedt et al. (2011:52) conclude that "In summary, our key finding and decision rule are to use AIC3 and CAIC jointly when evaluating FIMX-PLS results."

To sum up, current evidence from the simulation studies considered in this research suggest that the accuracy of commonly used model selection criteria for determining the number of segments in a sample strongly depends on the usage context, including the types of distributions used to describe the data, the model specification, and the characteristics of the specific market. However, these results also indicate that AIC3 seems to be a good criterion to use across a wide variety of model specifications and data configurations.

3.3.2 Variable selection

Steinley and Brusco (2008b) recently evaluated eight variable selection techniques for MB (Law, Figueiredo and Jain, 2004; Raftery and Dean, 2006; and Dy and Brodly, 2004) and NMB clustering. They used 20412 datasets, each one generated with 250 observations and systematically manipulated factors, such as the number of segments, true structure variables, and masking variables, as well as the density of the segments, the average probability of overlap between segments on each true structure variable and the degree of within-segment correlation. The most effective method was the procedure proposed by Steinley and Brusco (2008a) for *k*-means. They found that variable selection methods used in conjunction with FMMs performed the worst suggesting that a business analyst should avoid using or use variable selection methods with FMMs only if it is necessary to do so.

3.4 Other important issues

3.4.1 Stability and validity

Brun et al. (2007) investigated the performance of internal (trace criterion, determinant criterion, invariant criterion, correlation with Euclidean distance matrix, silhouette index), relative (figure of merit, stability) and external (Hubert's correlation, Rand statistic, Jaccard coefficient, Folkes and Mallows index) validation indices[3] applied to the outcomes of several clustering algorithms (*k*-means, fuzzy *c*-means, self organizing maps (SOM), single, complete and average linkages, and MB clustering methods) under realistic conditions in order to evaluate their performances. They used several models with different mixtures (regarding dimensionality and shape of the mixture distributions considered). They conclude that the Rand statistic and the silhouette index are the best performing external and internal validation indices, respectively. A business researcher should therefore consider these measures for external and internal validation, respectively.

3.4.2 Software packages

Haughton, Legrand and Wolford (2009) recently compared three software packages, Latent GOLD, MCLUST and poLCA that can be used to perform LC-based market segmentation. Using a dataset having continuous, discrete or mixed variables and by applying each software package to develop a LC CA for this data, they were able to compare software features and the resulting clusters. The results obtained using MCLUST outperformed those by Latent GOLD according to the measure of heterogeneity they used. From the perspective of usability they concluded that Latent GOLD is the easiest to use with a well-written and usable documentation and a GUI interface that eliminates the need for user programming.

4. Methodology of this study

4.1 Instrument

The study instrument used to gather data for this research was developed based on a comprehensive review of the FMM literature. In addition to general information such as authorship and publishing data, the coded criteria reflect the basic structure of a typical FMM application in MS. These criteria can be divided into four broad stages representing the number of steps that are important to the quality of a segment solution. Table 5 specifies the data collected for the literature analysis.

Table 5: Data collected in the literature analysis

Criteria		Data collected
Model type		Type of mixture model used as stated by the authors
Parameter estimation		Methods used for parameter estimation (e.g. ML or Bayesian methods, initialization methods, convergence, identifiability, number of iterations, type of distribution, label switching)
Model selection	Determination of the number of clusters, variable selection and related issues	Number of segments; model selection criteria and reasons for their usage; segment sizes; segmentation variables; number of variables used in the segmentation; sample size
	Data pre-processing	Data pre-processed before being clustered; data pre-processing method; number of variables before and after data pre-processing; reasons for data pre-processing
Other important issues	Validity and stability	Evaluation of stability and validity; methods used
	Interpretation and description of segments	No description; partial description; full description
	Software used (FMM-related packages)	Type of software (e.g. Latent Gold or MCLUST)

[3] Please see Brun et al. (2007) for a detail description of the indices.

4.2 Data collection and analysis

As the nature of research on FMM applications is difficult to confine to specific disciplines, the relevant material is scattered across various journals. We therefore searched online journal databases to obtain a representative as possible bibliography of FMM applications in marketing and business literature. In doing so, most of the top-tier marketing, business, management and tourism journals were included in the literature analysis. The literature search was based on numerous descriptors, such as 'mixture model(s)', 'finite mixture model(s)', 'latent class model(s)' (LCMs), 'latent class analysis', 'multilevel latent class analysis', 'finite mixture structural equation model(s)', 'finite mixture partial least squares', 'Bayesian mixture model(s)' (BMM), 'latent class', 'mixture model', 'hidden Markov models' (HMM), etc. in conjunction with the words 'segmentation', 'marketing' and 'business'.

The full text of each article was reviewed to eliminate those articles that were not related to marketing/business and to the objectives of the study. The selection criteria were as follows: Only those FMM application articles that had been published in journals within the target timeframe (from 2000) in a marketing setting were selected. This search yielded 108 articles from 63 journals with about 40 percent of the articles coming from at least category A journals according to the 'Journal Quality List' of Harzing (2011).

5. Results

5.1 Finite mixture models used

An analysis of the data by the type of mixture models used as stated by the respective authors unveiled that FMMs and LCMs are dominant in MS. The other cited mixture model approaches – Bayesian methods, FIMIX-PLS, FIMIX-SEM, HMM, MLC with 2.8, 1.9, 1.9, 4.6 and 1.9 percent of applications, respectively – seemingly did not enjoy the same popularity as FMMs (29.6%) and LCMs (51.9%) in the considered time-period.[4] 4.6% of the studies used latent class mixture models. In one study the name of the software package was stated as the method used. In 18.5, 9.3 and 1.8 percent of the studies regression, logit and probit models were used respectively. Compared to Wedel and Kamakura (2000), these results show a considerable increase in the usage of new FMM types.

5.2 Parameter estimation in mixture models

ML (41.7%) is the most popular method used for parameter estimation. Bayesian methods were applied in 6.5% of the studies and, remarkably, in 51.9% the method used was not ascertainable. Despite some of its known shortcomings, the EM algorithm, used in 17.6% of the studies, was the most common method for parameter estimation. Numerical methods were used in 8.3% of the studies considered and in one study EM and MCMC were combined. Although MCMC clearly outperformed the EM algorithm in the study of Dias and Wedel (2004), it was implemented only in 6.5% of the studies considered.

The normal (Gaussian) distribution was used in 13.9% of the studies. Other types of distributions used include multivariate normal (9.3%), multinomial (5.6%), Poisson (3.7%), binomial (1.9%), Dirichlet-multinomial (1.9%), logistic (0.9%), Laplace (0.9%), Wishart (0.9%) and truncated bivariate normal (0.9%). Normal and multinomial distributions were combined in two studies (1.9%) and normal and logistic in one (0.9%). The distribution was not stated in 57.4% of studies.

Convergence was assessed in only 13.9% of the studies. The most popular method used to assess convergence was examining the iteration plots, followed by Geweke's (1992) convergence test (Rust and Verhoef, 2005). Other methods used to assess convergence include comparing the within to between variance for each parameter estimated across multiple chains (Netzer, Lattin and Srinivasan, 2008) and Heidelberger and Welch stationarity test (Rust and Verhoef, 2005).

Only 13% of the studies used different starting values to initialize the algorithm. Random starting points were used by almost all studies. Only 19.4% of the studies used several runs of the algorithm to overcome the problem of local maxima. Identifiability and label switching were addressed in 16.7% and 4.6% of the studies, respectively.

[4] In this context, it has to be mentioned that both terms, FMM and LCM, are often used in a synonymous way.

These results show that crucial factors that can be assumed to significantly impact the final solution are not adequately addressed in many of the studies – estimation methods used (51.9%), convergence of the algorithm used (86.1%), initializing the algorithm (87%), several runs of the algorithm (80.6%), identifiability (83.3%) and label switching (95.4%).

5.3 Model selection

5.3.1 Methods used to determine the number of segments

Figure 1 shows the methods used to determine an optimal/adequate number of segments in FMM applications. The category "Others" include those methods which were used only once. 12% of the studies did not indicate how they arrived at the number of segments. Interestingly, despite the repeated successes of AIC3 in determining the true number of segments in many simulation studies, its usage seemingly is not widespread in empirical applications. An investigation of the reasons for selecting a particular model selection criterion reveals that 87% of studies did not put forward any reasons for using a particular procedure to determine the number of segments.

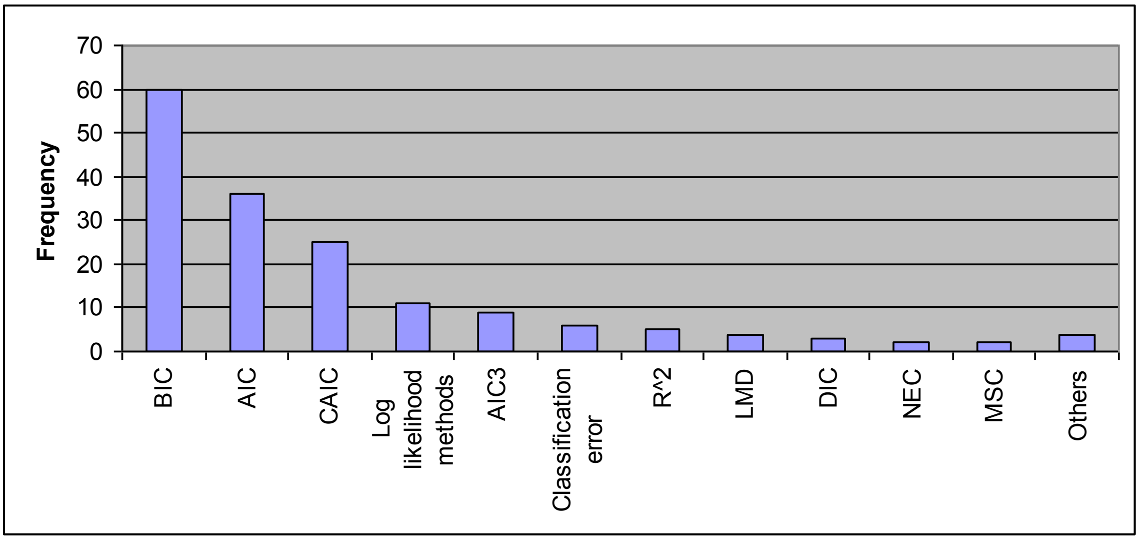

Figure 1: Methods used to determine the number of segments

Confidence in the number of clusters/segments is greater when multiple methods used to determine them converge. However, only 37% of the studies used this approach.

The minimum number of segments derived was two and the maximum was 14. On average 3.5 segments were derived. The number of segments was not ascertainable in 5.6% of the studies. Atella et al. (2004:660) state that in applications of FMMs most studies find that only a small number of latent segments are needed to describe the data adequately. Using additional data from Tuma, Decker and Scholz (2011) a one-way ANOVA test was carried out to check whether the clustering methods used (hierarchical, non-hierarchical methods and FMMs) have an influence on the number of segments derived. The test results ($p < 0.01$) are highly significant suggesting that FMMs maybe the appropriate method to use when the analyst suspects the existence of only a few segments in a dataset. The segment sizes were ascertainable in only 70.4% of the studies. This is a cause for concern considering that many managerial decisions are based on this information. The smallest segment used in the reviewed studies contains no more than seven respondents. The maximum segment size had 334083 respondents. 6.3% of the studies did not comply with the rule of thumb suggested in Garver, Williams and Taylor (2008).

5.3.2 Variable selection, number of variables used in clustering and sample size

The outcome of an analysis of the data regarding variable selection is illustrated in Figure 2. As the bar chart shows, there is some variety in the use of segmentation variables in FMM applications. In approximately 5% of the articles no information is provided about the segmentation variables used. Given that the segmentation variables highly impact the resulting segmentation solution, this finding is a cause for concern. Despite the criticism of demographic variables, they continued to be used intensively in segmentation studies.

A promising new source for obtaining interesting segmentation variables is online consumer reviews as provided for example by Epinions.com. As Decker and Trusov (2010) point out, the analysis of freely expressed customer opinions is a promising alternative to conventional survey techniques since the reviewers or respondents have not been requested to communicate their opinion, but are doing this voluntarily. Accordingly, a high level of authenticity can be expected. None of the articles considered used this promising data source for segmentation variable elicitation. 70% of the segmentation data comes from opinion surveys (face-to-face interviews, mail interviews, online surveys, etc). 25.5% come from databases and the remaining 4.5% is unascertainable.

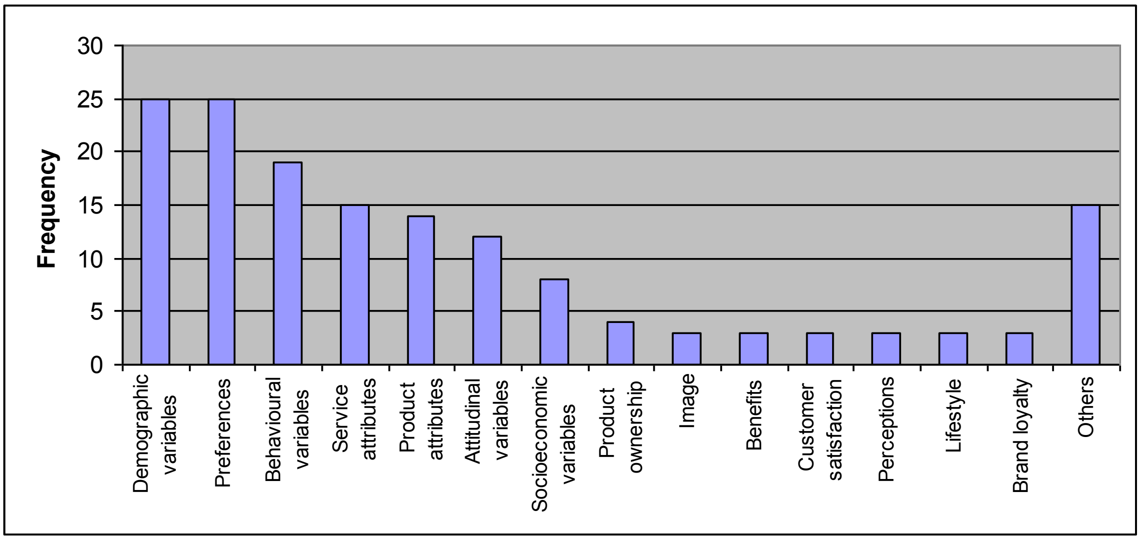

Figure 2: Segmentation variables used in FMM applications

The data were also analysed by looking at the number of variables used, the sample sizes and the relationship between these. The sample sizes were ascertainable in 97.2% of the studies and the numbers of segmentation variables was made available in 86% of the papers considered. A noteworthy 60% of the studies included in this literature analysis did not comply with the rule of thumb suggested by Formann (1984).

In 22.2% of the studies, the data were pre-processed, reducing the number of variables on average from 25 to 6 equalling a reduction of 76%. 9.3% of the studies included in this review factor analyzed the segmentation variables. The variables were standardized in an equal amount of studies. In 3.7% of the studies standardization and factor analysis were combined.

5.4 Other important issues

Validity and stability may be the most neglected issues in segmentation analysis. Surprisingly, they were not investigated in 83.3% and 65.7% of the studies respectively. The validation methods used include cross-validation (hold-out) sampling (26.9%), external validation (5.6%) and internal validation (1%). A combination of all three methods was reported in one study.

The segments were not described or interpreted in 16.7% of the studies. This is of some concern considering that the description and interpretation of segment solutions are the basis for formulating credible positioning and targeting strategies, for example.

In 64.8% of all FMM publications, the software used was not stated. Latent Gold, Gauss, LIMDEP/NLOGIT, Mplus, SmartPLS, LEM Software and SAS were used in 8.3, 6.5, 4.6, 3.7, 2, 2, and 2% of the studies, respectively. MCLUST, WinBUGS, Stata, poLCA and EMMIX were used in one study each. In one study Mplus and LIMPDEP\NLOGIT were combined.

6. Conclusion and outlook

MS remains the bedrock on which the positioning and targeting decisions of many businesses are built. Obtaining robust and valid solutions is therefore of crucial importance. FMMs have been promoted by some methodologist as a sophisticated method for obtaining good segmentation solutions.

In this study, we complement the toolbox of business practitioners by presenting FMMs as a comparatively new and interesting class of business research methods for deriving homogeneous groups of customers, for example. The presentation of set-ups and key results of simulation studies was intended to provide guidance for those seeking an appropriate FMM, model selection criteria, validation procedures, etc. for similar segmentation problems. Furthermore, identifying FMMs and methods for model selection, parameter estimation, variable selection and validation that are most robust to variations in data characteristics is of great importance for business researchers focusing on data-based market segmentation since they can use these methods with more confidence regarding the accurate recovering of consumer behaviour and characteristics. The presented results, among others, show that AIC3, MCMC, the absolute convergence criterion in conjunction with the EM algorithm and RP with the SEM, and MCLUST are useful methods or tools across a wide variety of model specifications and data configurations.

In order to review the common practices of marketing researchers, we carried out a comprehensive survey of 108 articles in which FMMs were applied in MS. The literature analysis was intended to serve as an orientation for marketing researchers interested in both best practices and pitfalls. The results show that crucial factors such as estimation methods, initializing and convergence of the algorithm used, identifiability and label switching, model selection criteria, stability and validity, etc. which may impact the final segmentation solution significantly are not always adequately addressed and reported in detail. Failure to provide specific information about the segmentation variables or method(s) used tends to inhibit replication and provides little guidance for researchers or business practitioners who might seek an appropriate method for a similar problem. Furthermore, the results show a remarkable discrepancy between simulation studies and applied FMMs methods. Methods or tools that performed well in simulation studies such as AIC3, MCMC and MCLUST are hardly used in practice.

Based on the available results, we believe that several issues have to be addressed by business researchers. Firstly, they should attempt to integrate best practice recommendations into their use of FMMs. Methods that performed well in simulation studies should be considered more intensively in empirical FMM applications. Furthermore, academic researchers should pay more attention to a detailed description of the FMM method used in their publications in order to provide better guidance to those who want to apply the respective (new) approach in practice. And last but not least, the idea of finite mixture modelling should increasingly find its way into education in business and marketing management in order to enable future generations of business and marketing researchers to successfully apply this powerful new class of market segmentation methods.

Future research should focus on large-scale simulation studies using a wide range of models and statistical distributions, and also investigate in-depth the effects of distributional misspecifications on the segmentation results. Furthermore, promising new data sources such as online consumer reviews should be considered in model-based segmentation. Model-based reviewer clustering, for instance, can help to better understand which reviewer characteristics have an influence on trust in this emerging source of word-of-mouth.

References

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

Andrews, R. L., Ainslie, A. and Currim, I. S. (2002) 'An empirical comparison of logit choice models with discrete vs. continuous representations of heterogeneity'. *Journal of Marketing Research*, 39: 479-487.

Andrews, R. L., Brusco, M. J. and Currim, I. S. (2010) 'Amalgamation of partitions from multiple segmentation bases: A comparison of non-model-based and model-based methods'. *European Journal of Operational Research*, 201: 608-618.

Andrews, R. L., Brusco, M. J., Currim, I. S., and Davis, B. (2010) 'An empirical comparison of methods for clustering problems: Are there benefits from having a statistical model?', *Review of Marketing Science*, 8: Article 3.

Andrews, R. L. and Currim, I. S. (2003a) 'A comparison of segment retention criteria for finite mixture logit models'. *Journal of Marketing Research*, 40 (2): 235-243.

Andrews, R. L. and Currim, I. S. (2003b) 'Retention of latent segments in regression-based marketing models'. *International Journal of Research in Marketing*, 20 (4): 315-321.

Andrews, R. L., Currim, I. S., and Leeflang, P. (2011) 'A comparison of sales response predictions from demand models applied to store-level versus panel-level data'. *Journal of Business and Economic Statistics*, 29 (2): 319-326.

Atella, V., Brindisi, F., Deb, P. and Rosati, F. C. (2004) 'Determinants of access to physician services in Italy: A latent class seemingly unrelated probit approach'. *Health Economics*, 13: 657-668.

Bekker, P., Merckens, A., and Wansbeek, T. J. (1994) *Identification, equivalent models and computer algebra*, New York: Academic Press.

Blashfield, R. K. (1976) 'Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods', *Psychological Bulletin,* 83: 377-388.

Bozdogan, H. (1987) 'Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions'. *Psychometrika*, 52 (3): 345-370.

Bozdogan, H. (1994) 'Mixture model cluster analysis using selection criteria and a new informational measure of complexity, in Bozdogan, H. (Ed.) *Multivariate Statistical Modelling*, Dordrecht: Kluwer Academic Publishers, 2: 69-113.

Brun, M., Sima, C., Hua, J., Lowey, J., Carroll, B., Suh, E. and Dougherty, E. R. (2007) 'Model-based evaluation of clustering validation measures'. *Pattern Recognition*, 40: 807-824.

Celeux G. (1998) 'Bayesian inference for mixture: The label switching problem' in Payne R. and Green P. (Eds.), *COMPSTAT 98*, Heidelberg: Physica, 227-232.

Celeux, G. and Soromenho, G. (1996) 'An entropy criterion for assessing the number of clusters in a mixture model'. *Journal of Classification,* 13: 195-212.

Celeux, G., Hurn, M., and Robert, C. P. (2000) 'Computational and inferential difficulties with mixture posterior distributions'. *Journal of the American Statistical Association*, 95: 957–970.

Decker, R. and Trusov, M. (2010) 'Estimating aggregate consumer preferences from online product reviews'. *International Journal of Research in Marketing*, 27 (4): 293-307.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) 'Maximum likelihood from incomplete data via the EM algorithm'. *Journal of the Royal Statistical Society Series B (Methodological)*, 39 (1): 1-38.

Dias, J. (2006) 'Model selection for the binary latent class model: A Monte Carlo simulation' in Batagelj, V., Bock, H.-H., Ferligoj, A., Žiberna, A. (Eds.), *Data Science and Classification*, Berlin: Springer, 91-99.

Dias, J. and Wedel, M. (2004) 'An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods'. *Statistics and Computing*, 14: 323-332.

Dy, J. G. and Brodley, C. E. (2004) 'Feature selection for unsupervised learning'. *Journal of Machine Learning Research*, 5: 845-889.

Ebbes, P., Grewal, R. and DeSarbo, W. S. (2010) 'Modeling strategic group dynamics: A hidden Markov approach'. *Quantive Marketing and Economics,* 8: 241-274.

Formann, A. K. (1984) *Die Latent-Class-Analysis: Eine Einführung in die Theorie und Anwendung*, Weinheim: Beltz.

Frühwirth-Schnatter, S. (2006) *Finite Mixture and Markov Switching Models*, New York: Springer.

Garver, M. S., Williams, Z. and Taylor, G. S. (2008) 'Employing latent class regression analysis to examine logistics theory: An application to truck driver retention'. *Journal of Business Logistics*, 29 (2): 233- 257.

Hahn, C., Johnson, M. D., Hermann, A. and Huber, F. (2002) 'Capturing consumer heterogeneity using a finite mixture PLS approach'. *Schmalenbach Business Review*, 54 (3): 243-269.

Hair, J., Ralph. F., Anderson, E., Tatham, R. L. and William C. Black (1998) *Multivariate Data Analysis,* Englewood Cliffs: Prentice Hall.

Haughton, D., Legrand, P., and Wolford, S. (2009) 'Review of three latent class cluster analysis packages: Latent GOLD, poLCA, MCLUST'. *The American Statistician*, 63 (1): 81-91.

Harzing, A.-W. (2011) *Journal Quality List*, University of Melbourne, available at: www.harzing.com

Henson, J. M., Reise, S. P., Kim, K. H. (2007) 'Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics'. *Structural Equation Modeling*, 14 (2): 202-226.

Hofstede, F. T., Wedel, M., and Steenkamp, J.-B. E.M. (2002) 'Identifying spatial segments in international markets'. *Marketing Science*, 21 (2): 160-177.

Jedidi, K., Jagpal, H. S. and DeSarbo, W. S. (1997) 'Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity'. *Marketing Science*, 16 (1): 39-59.

Law, M. H. C., Figueiredo, M. A. T. and Jain, A. K (2004) 'Simultaneous feature selection and clustering using mixture models'. *IEEE Transactions on Pattern Analysis and Machine Learning*, 26: 1154-1166.

Lubke, G. H. and Muthén, B. (2005) 'Investigating population heterogeneity with factor mixture models'. *Psychological Methods*, 10 (1): 21-39.

Lukočiene, O., Varriale, R., and Vermunt, J. K. (2010) 'The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis'. *Sociological Methodology*, 40 (1): 247-283.

Lukočiene, O. and Vermunt, J. K. (2010) 'Determining the number of components in mixture models for hierarchical data' in Fink, A., Lausen, B., Seidel, W., Ultsch, A. (Eds.), *Advances in Data Analysis, Data Handling and Business Intelligence*, Berlin: Springer, 241-249.

McLachlan, G. and Peel, D. (2000) *Finite Mixture Models*. New York: Wiley.

Melnykov, V. and Maitra, R. (2010) 'Finite mixture models and model-based clustering'. *Statistic Surveys*, 4: 80-116.

Netzer, O., Lattin, J. M. and Srinivasan, V. (2008) 'A hidden Markov model of customer relationship dynamics'. *Marketing Science*, 27 (2): 185–204.

Raftery, A. E. and Dean, N. (2006) 'Variable selection for model-based clustering'. *Journal of the American Statistical Association*, 101: 294-304.

Rust, R. T. and Verhoef, P. C. (2005) 'Optimizing the marketing interventions mix in intermediate-term CRM'. *Marketing Science*, 24 (3): 477-489.

Sarstedt, M. (2008) 'Market segmentation with mixture regression models: Understanding measures that guide model selection'. *Journal of Targeting, Measurement and Analysis for Marketing*, 16: 228-246.

Sarstedt, M., Becker, J.-M, Ringle, C. M. and Schwaiger, M. (2011) 'Uncovering and treating unobserved heterogeneity with FIMIX-PLS: Which model selection criterion provides an appropriate number of segments?'. *Schmalenbach Business Review*, 63 (1): 34-62.

Sarstedt, M. and Schwaiger, M. (2007) 'Model selection in mixture regeression analysis – A Monte Carlo simulation study', in Preisach, C., Burkhardt, H., Schmidt-Thieme, L. and Decker, R. (Eds), *Data Analysis, Machine Learning, and Applications*, Berlin: Springer, 61-68.

Schwarz, G. (1978) 'Estimating the dimension of a model'. *The Annals of Statistics*, 6 (2): 461-464.

Sclove, S. L. (1987) 'Application of model-selection criteria to some problems in multivariate-analysis'. *Psychometrika*, 52: 333-43.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002) 'Bayesian measures of model complexity and fit'. *Journal of the Royal Statistical Society, Series B*, 64, 583-640.

Steinley, D., and Brusco, M. J. (2008a) 'A new variable weighting and selection procedure for *k*-means cluster analysis'. *Multivariate Behavioral Research,* 43: 77-108.

Steinley, D., and Brusco, M. J. (2008b) 'Selection of variables in cluster analysis: An empirical comparison of eight procedures'. *Psychometrika,* 73: 125-144.

Stephens, M. (1997) *Bayesian Methods for Mixtures of Normal Distributions*, Dissertation, University of Oxford.

Tuma, M. N., Decker, R. and Scholz, S. W. (2011) 'A survey of the challenges and pitfalls of cluster analysis application in market segmentation'. *International Journal of Market Research*, 53 (3): 391-414.

Wedel, M. and DeSarbo, W. S. (2002) 'Market segment derivation and profiling via a finite mixture model framework'. *Marketing Letters*, 13 (1): 17-25.

Wedel, M. and Kamakura, W. A. (2000) *Market Segmentation: Conceptual and Methodological Foundations*, Dordrecht: Kluwer.