Market segmentation with mixture regression models: Understanding measures that guide model selection

Received (in revised form): 6th May, 2008

Marko Sarstedt

is an assistant professor at the Institute for Market-based Management at the Ludwig-Maximilians-University Munich. His research interests include research methodology, especially in the fields of finite mixture modelling and partial least-squares path analysis.

Keywords *information criteria, finite mixture models, unobserved heterogeneity, mixture regression, FIMIX-PLS, customer satisfaction*

Abstract Owing to their considerable potential for market segmentation studies, mixture regression models have recently received increasing attention from both academics and practitioners. One fundamental difficulty with their application is related to the problem of model selection, that is, the choice regarding the number of segments. Retaining the correct number of segments is, however, crucial as many managerial decisions depend on this decision. Since the proper number of segments is unknown in real-world applications, a thorough understanding of measures that guide the model selection decision is of fundamental importance. Based on a simulation study, this paper addresses the issue by evaluating how the interaction of the most important influencing factors for the measures' success — sample and segment size — affects the performance of four of the most widely used criteria for assessing the correct number of segments in mixture regression models. For the first time, the quality of these criteria is evaluated with regard to a wide spectrum of possible constellations. Furthermore, relative and absolute performances are analysed in respect of outside criteria. Recommendations on criterion selection are thereafter deduced from the results when a certain sample size is given. These recommendations also help to establish the sample size that is needed in order to guarantee an accurate decision based on a specific criterion. An application based on customer satisfaction data illustrates the relevance of the findings. In conclusion, theoretical and managerial implications are provided.

Journal of Targeting, Measurement and Analysis for Marketing (2008) **16**, 228–246. doi:10.1057/jt.2008.9;
published online 7 July 2008

INTRODUCTION

The application of regression-based marketing models — still the most common analysis procedure in this field — is usually based on the assumption that the analysed data originate from

a single population, that is, a unique global model represents all the observations well. In many real-world applications however, this assumption of homogeneity is unrealistic, as individuals are likely to be heterogeneous in their perceptions and evaluations of marketing constructs. For example, in a customer satisfaction analysis, consumers may form different segments, each with different drivers of satisfaction.

Traditionally, heterogeneity in regression models is taken into account by assuming that observations can be assigned to segments *a priori*

Correspondence: Marko Sarstedt, Institute for Market-based Management, Ludwig-Maximilians-University Munich, Kaulbachstr. 45, Munich 80539, Germany.
Tel: +49 89 2180 5634;
Fax: +49 89 2180 5651;
E-mail: sarstedt@bwl.lmu.de

on the basis of, for example, psychographic or demographic variables.¹ In the case of a customer satisfaction analysis, this may be achieved by distinguishing high- and low-income user segments and carrying out separate analyses for each segment. Modelling segments based on *a priori* information, however, suffers from serious limitations. In many instances, substantive theory on the variables causing heterogeneity is not available or incomplete. Furthermore, observable characteristics such as gender, age or usage frequency are usually insufficient to capture heterogeneity adequately.² In other words, heterogeneity is often unobservable and the true causes hidden. Alternatively, by applying a cluster analysis, a researcher can partition the sample into segments and consecutively carry out segment-specific analyses. Different clustering algorithms, however, yield different results and to date there is little guidance on choosing the best procedure.

In the last years, another class of procedures, finite mixture models, have received increasing attention from both a practical and theoretical point of view. A finite mixture approach to model-based clustering assumes that the data originate from several subpopulations or segments.³ Each segment is modelled separately and the overall population is a mixture of segment-specific density functions. Consequently, homogeneity is no longer defined in terms of a common set of true scores, but at a distributional level. Thus, finite mixture modelling enables marketers to cope with heterogeneity in data by clustering observations and estimating parameters simultaneously, thus avoiding well-known biases that occur when models are evaluated separately.⁴ Correspondingly, mixture regression models are prevalent in marketing literature.^{5–8} For example, in a recent study, Andrews *et al.*⁹ use a mixture regression model approach to account for unobserved heterogeneity in the well-known SCAN*PRO model. Through this fit and prediction accuracy have been increased, compared to an aggregate-level analysis.

One of the most important mixture regression-based methodologies was presented by Hahn *et al.*¹⁰ and later advanced by Ringle *et al.*^{11,12} Their finite mixture partial least squares

(FIMIX-PLS) path modelling approach combines the advantages of finite mixture models with the strengths of the PLS method. Besides covariance-based structural equation modelling, PLS path modelling has established itself as another approach for estimating cause–effect relations between latent variables, based on a theoretically founded hypothesis system. FIMIX-PLS has been successfully applied in different context, such as customer satisfaction,^{13,14} brand preference^{15,16} and strategic success factors.^{17,18} In a recent contribution on the effects of corporate-level marketing activities on corporate reputation, Sarstedt¹⁹ applies the FIMIX-PLS procedure to disclose unobserved heterogeneity. The results demonstrate that the interpretation of PLS path modelling results in respect of the aggregate data level can be seriously misleading, thus causing erroneous inferences. Whereas the integration of *a priori* information has proved valuable in increasing model fit, the results could be improved by using FIMIX-PLS, which reliably identifies unobserved heterogeneity. In the light of these findings, Sarstedt *et al.*²⁰ conclude that ‘FIMIX-PLS will assume an imperative role in enhancing PLS in the next wave of analytical procedures’, which underlines the growing importance of the mixture regression concept. This notion is supported by a recent research which suggests that also finite mixture conjoint models achieve good parameter estimates, even at an individual level.²¹

One fundamental difficulty in the application of finite mixture models is related to the problem of model selection, that is, the decision regarding how many segments to retain from the data. A misspecification results in an under or oversegmentation, which may misguide marketing managers when deciding on, for example, customer targeting, product positioning or the determination of the optimal marketing mix.^{22,23} If the number of segments is over-specified, marketers may run the risk of treating audience segments separately even though they could be handled together more effectively. On the other hand, if a market is undersegmented, marketers may overlook distinct segments that could be addressed separately for more precise satisfaction of the customers’ varying wants.

Various authors have considered the problem of choosing the number of segments in mixture models in different context.^{24–27} As most of the available studies have appeared in statistics literature, they, however, aim at exemplifying the effectiveness of newly proposed measures, instead of revealing the performance of measures commonly available in statistical software programs. Furthermore, this topic has not been thoroughly considered for mixture regression models, despite their importance for the marketing research field. Exceptions in this area are presented by Hawkins *et al.*,²⁸ Andrews and Currim²⁹ and Oliveira-Brochado and Martins,³⁰ who examine the performance of various model selection criteria with regard to several factors such as measurement level and number of predictors, degree of separation between the segments or the error variance.

The following study argues that sample and segment size are the most critical influencing factors for the criteria's performance from both a practical as well as theoretical point of view. The interaction of these factors has hitherto not been investigated thoroughly. This paper aims at filling this gap in research by evaluating how the interaction of sample and segment size affects the performance of four of the most widely used criteria for assessing the number of segments in mixture regression models in detail. Relevant criteria were identified through a meta-study of papers that appeared in major marketing journals between 2000 and 2006. These were then evaluated by conducting a Monte Carlo simulation for a two-segment solution where the sample size was varied in a ten-step interval of [50;500]. For each sample size, five variations of relative segment sizes were evaluated. In order to assess the performance of each criterion, the rates at which the right number of segments were successfully chosen were computed.

Another shortcoming of existing studies is their mere focus on the criteria's relative effectiveness, thus ignoring any *a priori* information on the likelihood of a certain model occurring. Consequently, the success rates of the simulation were compared with an outside criterion – chance models. These chance models are derived from

discriminant analysis and are used to evaluate the information criteria's absolute performance with respect to chance. This provides researchers and practitioners with a better understanding of these criteria's effectiveness.

The rest of the paper proceeds as follows. The next section describes the theoretical background of mixture regression models and provides a general review of model selection criteria. In order to identify commonly used information criteria, a meta-study was conducted on the utilisation of statistical measures for model selection, the results of which are presented in this section as well. Thereafter, the design is introduced, which is followed by the simulation study's results. An empirical application of mixture regression models using customer satisfaction data underlines the relevance of the findings from a managerial standpoint. Finally, the study's key contributions as well as managerial implications and suggestions for further research are presented.

THEORETICAL BACKGROUND

Mixture regression models

A mixture model-based approach to regression analysis assumes that the observations of a data set originate from various groups with unknown segment affiliations. This heterogeneity is treated in simultaneous equation models by deriving segments that are homogenous with respect to the model's predictor values. That is, each observation is taken to be a realisation of the unconditional density

$$f(y_n | \varphi) = \sum_{s=1}^S \pi_s f_s(y_n | \theta_s) \quad (1)$$

with y_n as the dependent variable ($n=1, \dots, N$), π_s as the relative size (mixture proportion) of segment s ($\sum_{s=1}^S \pi_s = 1$ and $\pi_s > 0 \forall s = 1, \dots, S$) and $\varphi = (\pi_1, \dots, \pi_S; \theta_1', \dots, \theta_S')$ as the vector of all unknown parameters associated with the density function, where θ_s is the segment-specific parameter vector for the density function. Equation (1) describes a mixture linear regression (also latent class regression or cluster-wise regression) if the conditional density function

f_s is a normal density with the segment-specific mean $\beta_s'x$ and variance σ_s^2 .

Applications of mixture regression models are typically classified according to the distribution of the dependent variable.³¹ The most important distributions are normal, gamma or exponential for continuous variables, and binomial, multinomial or Poisson for discrete variables. As all these distribution types are within the exponential family, generalised linear models, including linear regression, logit or probit models can be applied.

Model selection in mixture models

Assessing the number of segments in a mixture model is a difficult but important problem.

Whereas it is well known that conventional χ^2 -based goodness-of-fit tests and likelihood ratio tests are unsuitable to determine the number of segments,^{32–34} the question regarding what model selection statistic should be used, remains unresolved.^{35,36} A modified likelihood ratio test uses a bootstrapping procedure to circumnavigate the implementation problems of classical χ^2 -tests.³⁷ This approach, however, requires vast computing power³⁸ and therefore lacks general application. The more recently developed Lo-Mendell-Rubin test compares two neighbouring models and provides a p -value to contrast the increase in model fit between $S-1$ and S segment models by using an approximate reference distribution for the log likelihood difference.³⁹ This method has nevertheless been criticised by Jeffries⁴⁰ due to its analytic inconsistency, which calls the validity of testing non-nested models with this method into questions.

The other main approach for deciding on the number of segments is based on a penalised form of likelihood, yielding the so-called information criteria. Information criteria for model selection simultaneously take into account the goodness-of-fit (likelihood) of a model and the number of parameters used to achieve that fit. The information criteria therefore denote a penalised likelihood function, that is, the negative likelihood plus a penalty term, which increases with the number of parameters and/or the number of observations.

Information criteria generally take the following form

$$-2\ln L + am(s) + b \quad (2)$$

where $\ln L$ denotes the maximum of likelihood over the parameters and $m(s)$ is the number of independent parameters in a model with s segments. For a given criterion, a is the cost of fitting an additional segment and b is an auxiliary term depending on the criterion. According to these criteria, in a set of competing models, the model that minimises the value in equation (2) should be chosen.

Various information criteria have been developed in recent years, whereat this study focuses only on four of the most representative and widely applied measures in marketing practice.

Commonly used information criteria in marketing studies

A thorough literature review revealed that only Oliveira-Brochado and Martins⁴¹ evaluated the importance of various information criteria in marketing studies. In their study, the authors report that in 37 published studies, Akaike's information criterion (AIC)⁴² was used 15 times, consistent AIC (CAIC)⁴³ was used 13 times and Bayesian information criterion (BIC)⁴⁴ was used 11 times (multiple selections possible).

Since it is unclear in which publications the studies appeared, a meta-study was initiated to identify the most commonly used information criteria in the field of marketing. For this purpose, all marketing journals were considered that are rated *A* or *A+* in the rankings developed on behalf of the Vienna University of Economics and Business Administration in 2001 and the Association of University Professors of Management in German-speaking countries (VHB) in 2003 (for a complete list, see Journal Quality List by Harzing⁴⁵). In order to make the data construction as transparent as possible, an easily accessible but universal research database was used. In November 2006, the EBSCO database was searched for any reference to 'finite mixture', 'mixture regression', 'cluster-wise regression' and 'latent class regression'. EBSCO is

the most comprehensive full text database for peer-reviewed research papers. The EBSCO search led to a great number of references of which a large fraction covered an entirely different topic. The empirical papers were examined with respect to whether they actually used any type of mixture regression analysis. Eventually, the desired results could be gained from 33 articles that appeared between January 2000 and November 2006. Table 1 presents the results of the meta-study.

The paper by Ho and Chong⁷⁹ was omitted from the analysis due to missing specifications on the model selection statistic used. In the remaining 32 articles, the problem of model selection and the decision regarding an appropriate statistical figure are only addressed four times: Danaher and Mawhinney,⁸⁰ Danaher⁸¹ as well as Wu and Rangaswamy⁸² refer to a contribution by Bucklin and Gupta,⁸³ who apply

multiple-segment choice models to capture customer heterogeneity in brand choice. In this study, the authors briefly discuss the applicability of likelihood ratio tests, the AIC and the BIC from a theoretical point of view without referring to any simulation study results. Only Andrews *et al.*⁸⁴ justify their choice of a model selection criterion. This choice is based on an – at that time unpublished – *Journal of Marketing Research* paper that presents the results of a simulation study comparing several criteria's performances.⁸⁵ In the remaining 28 studies, no rationale whatsoever is given for the model selection statistics chosen. In none of the studies did the authors refer to test statistics to decide on the number of segments in the mixture. In fact, all authors refer to information criteria for that decision.

In the studies, BIC was used 25 times, AIC was used eight times and CAIC as well as modified AIC with factor three (AIC₃)⁸⁶ was used twice (multiple selections possible). In six cases, more than one information criterion was applied. In four cases, the log-marginal density (LMD), which is computed as the logarithm of the harmonic means of the likelihood values, is used as an in-sample fit criterion. Since likelihood values are obtained by means of the estimated parameter samples drawn by Gibbs sampling, which is not applicable in this context, the LMD criterion is not considered in this study.

Most of these criteria have also been implemented in the two most popular commercial software programs for estimating mixture regression models – Latent Gold⁸⁷ and Mplus⁸⁸ – as well as in free software packages such as FlexMix.^{89,90}

Consequently, the following criteria were considered in the simulation study, which takes the following form:

$$AIC = -2\ln L + 2m(s) \quad (3)$$

$$AIC_3 = -2\ln L + 3m(s) \quad (4)$$

$$CAIC = -2\ln L + m(s)[\ln(N) + 1] \quad (5)$$

Table 1: Results of the meta-study

Year	References	Criterion/criteria considered
2000	Bell and Lattin ⁴⁶	BIC
2000	Heilman <i>et al.</i> ⁴⁷	BIC
2000	Mazumdar and Papatla ⁴⁸	BIC
2000	Shachar and Emerson ⁴⁹	AIC, BIC, CAIC
2001	Danaher and Mawhinney ⁵⁰	BIC
2001	Erdem <i>et al.</i> ⁵¹	BIC
2001	Gönül <i>et al.</i> ⁵²	AIC, BIC
2001	Thomas ⁵³	AIC
2002	Andrews <i>et al.</i> ⁵⁴	BIC
2002	Andrews <i>et al.</i> ⁵⁵	BIC, LMD
2002	Danaher ⁵⁶	BIC
2002	Hofstede <i>et al.</i> ⁵⁷	LMD
2002	Hofstede <i>et al.</i> ⁵⁸	LMD
2002	Papatla and Bhatnagar ⁵⁹	BIC
2002	Wedel and DeSarbo ⁶⁰	CAIC
2003	Agarwal ⁶¹	BIC
2003	Chung and Rao ⁶²	LMD
2003	Danaher <i>et al.</i> ⁶³	BIC
2003	Ho and Chong ⁶⁴	ns
2003	Wu and Rangaswamy ⁶⁵	AIC, BIC, AIC ₃
2004	Anand and Shachar ⁶⁶	BIC
2004	Bowman <i>et al.</i> ⁶⁷	AIC, BIC
2004	Lewis ⁶⁸	BIC
2004	Varki and Chintagunta ⁶⁹	BIC
2004	Zhang and Krishnamurthi ⁷⁰	BIC
2005	Jedidi and Kohli ⁷¹	BIC
2005	Lewis ⁷²	BIC
2005	Reinartz <i>et al.</i> ⁷³	AIC
2005	Rust and Verhoef ⁷⁴	AIC, BIC, AIC ₃
2005	Thomas and Sullivan ⁷⁵	AIC
2006	Kivetz <i>et al.</i> ⁷⁶	BIC
2006	Mantrala <i>et al.</i> ⁷⁷	BIC
2006	Srinivasan ⁷⁸	BIC

$$BIC = -2\ln L + \ln(N)m(s) \quad (6)$$

For a description of the theoretical underpinnings and statistical properties, compare the references cited above.

Empirical comparisons of criteria in mixture regression models

Despite the widespread use of mixture regression models and the importance of ascertaining the true number of segments in order to reach meaningful conclusions from any analysis, only three studies have so far observed the performance of information criteria in this methodological context. The study by Andrews and Currin⁹¹ examines the performance of AIC, AIC₃, CAIC, BIC, information complexity criterion (ICOMP),^{92,93} the validation sample log likelihood (LOGLV) and the normed entropy criterion (NEC)⁹⁴ by determining the success rates of each criterion under consideration of the root mean square error between the true and estimated parameters of the chosen model. The examination is based on simulated data sets with eight factors, which, according to the literature, potentially affect the criteria's performance. Under most experimental conditions, AIC₃ demonstrates the best overall performance, followed by LOGLV and BIC, the latter dominating CAIC. Lastly, AIC shows high overfitting rates and ICOMP low overall success rates. The authors conclude that AIC₃ is the best criterion to use with regression models for normally distributed data. Although the study provides good insight into the criteria's overall performance, it remains unclear in which factor level combination each criterion operates favourably or not.

In their simulation study, Hawkins *et al.*⁹⁵ consider similar information criteria to those used in the study by Andrews and Currin⁹⁶ and evaluate the influence of segment separation and mixing proportions. The authors conclude that BIC is the recommended criterion for choosing between one and two-segment mixtures of linear regression models. For three and four segments, none of the measures outperforms the others across the simulation runs, whereby for $S=4$, all criteria perform rather weakly, showing success

rates of less than 50 per cent. The authors, however, state that the simulation results are limited since the effects of small sample sizes were not explored.

In a more recent study, Oliveira-Brochado and Martins⁹⁷ examine the model selection criteria's performance in recovering small niche segments. Furthermore, the authors evaluate the impact of distributional misspecification of the error term. The experimental design comprises data sets with six predictors, an alternating number of segments and mean separations between segment coefficients. In the niche segment case, AIC and AIC₃ show the best performance, whereas BIC and CAIC, just like most of the other criteria not presented in this paper, achieve rather low success rates. Furthermore, segment retention criteria did not, after all, perform less well in situations with distributional misspecification where the error term followed a uniform distribution.

Despite the broad scope of questions covered in these studies, they do not profoundly investigate the criteria's performance in respect of the one factor best influenceable by the marketing analyst, namely the sample size: whereas Andrews and Currin⁹⁸ allow for only two levels of sample size ($N=100$ and 300), Hawkins *et al.*⁹⁹ as well as Oliveira-Brochado and Martins¹⁰⁰ consider just one sample size ($N=500$ and 300 , respectively). From an application-oriented point of view, it is, however, desirable to know which sample size is required in order to guarantee validity when choosing a model with a certain criterion. Furthermore, disregarding of the sample size proves problematic not only from a practical point of view, but also from a theoretical one. As indicated above, the sample size is a key differentiator between the criteria, because $-2\ln L$ from equation (2) remains constant for all criteria described in equations (3)–(6). Consequently, sample size must have a great effect on the criteria's effectiveness since the available studies yield different conclusions about the criteria's performance. Therefore, the first objective of this study is to determine how well the information criteria perform in a mixture regression of normal data with alternating sample

sizes, which answers the call for further research by Hawkins *et al.*¹⁰¹

Another factor closely related to this problem concerns the relative segment sizes. Even though a specific sample size might prove beneficial in guaranteeing a satisfactory performance of the information criteria in general, the presence of a niche segment might lead to a reduced heterogeneity and thus to a wrong decision in choosing the number of segments. Furthermore, the studies mentioned before indicated that the relative size of the segments significantly influences the criteria's success rates. Consequently, the second objective is to evaluate the influence of varying mixture proportions on the information criteria's performance.

The joint influence of these factors is evaluated across a broad range of levels for a two-segment solution.

SIMULATION DESIGN AND EVALUATION OF RESULTS

The simulation strategy consists of initially drawing observations derived from an ordinary least-squares regression and applying these to the FlexMix algorithm.^{102,103} FlexMix is a general framework for finite mixtures of regression models using the EM algorithm,¹⁰⁴ which is available as an extension package for the statistical computing software R.¹⁰⁵ R provides functionalities for a flexible design of computational experiments, which, together with the package's extensibility,¹⁰⁶ made FlexMix preferable to alternative commercial solutions.

In this simulation study, models with alternating numbers of observations $N_i (i = 1, \dots, 46)$ and three continuous predictors were considered for the ordinary least-squares regression. First, criterion variable scores were computed for each observation, with the predictor variable values drawn from a standard normal distribution. Subsequently, an error term derived from a standard normal distribution was added to the criterion variables' scores. For each simulation set-up (ie each combination of sample and segment size), 1,000 data sets were generated, totalling $46 \times 5 \times 1,000 = 230,000$ data sets. The

main parameters controlling the simulation were as follows:

- The number of segments: $S = 2$.
- The regression coefficients in each segment that were specified as follows:
 - Segment 1: $\beta_1 = (1, 1, 1.5, 2.5)'$
 - Segment 2: $\beta_2 = (1, 2.5, 1.5, 4)'$
- The sample sizes that were varied in a ten-step interval of [50;500].
- The size of the segments. For each sample size, the simulation was run for the following five mixture proportions: $(\pi_1^1 = 0.1, \pi_2^1 = 0.9)$; $(\pi_1^2 = 0.2, \pi_2^2 = 0.8)$; $(\pi_1^3 = 0.3, \pi_2^3 = 0.7)$; $(\pi_1^4 = 0.4, \pi_2^4 = 0.6)$; $(\pi_1^5 = 0.5, \pi_2^5 = 0.5)$.
- The range of potential solutions considered, whereby each simulation run was carried out five times for $S = 1, \dots, 5$ segments.

The likelihood was maximised using the EM algorithm. As a limitation of the algorithm is its convergence to local maxima,¹⁰⁷ it was run repeatedly with ten replications. Owing to the high number of data sets and the excessive computational demand, ten replications were deemed appropriate. The best solution was then chosen for each number of segments. The results were obtained by calculating the proportion of all the trials that were correct for each measure with respect to the sample size and mixture proportion.

As indicated above, previous studies have only observed the criteria's relative performance, ignoring the question whether the criteria perform any better than chance. To verify whether the criteria are adequate, the predictive accuracy of each criterion with respect to chance is measured using the following chance models derived from discriminant analysis: random chance, proportional chance and maximum chance criterion.¹⁰⁸

In order to be able to apply these criteria, the researcher has to have prior knowledge or make presumptions concerning the underlying model. For a given data set, let M_j be a model with S_j segments from a consideration set with C competing models $K = \{M_1, \dots, M_C\}$ and ρ_j be the prior probability to observe M_j ($j = 1, \dots, C$ and $\sum_{j=1}^C \rho_j = 1$ with $\rho_j > 0 \forall j = 1, \dots, C$).

The random chance criterion is

$$CM_{ran} = \frac{1}{C} = \rho \quad (7)$$

which indicates that each of the competing models has an equal prior probability.

The proportional chance criterion is

$$CM_{prop} = \sum_{j=1}^C \rho_j^2 \quad (8)$$

According to Morrison,¹⁰⁹ this criterion can mainly be used as a point of reference for subjective evaluation, rather than as the basis of a statistical test to determine whether the expected proportion differs from the observed proportion of models that is correctly classified.

The maximum chance criterion is

$$CM_{max} = \max(\rho_1, \dots, \rho_C) \quad (9)$$

which defines the maximum prior probability to observe model j in a given consideration set as the benchmark for a criterion's success rate. Since $CM_{ran} \leq CM_{prop} \leq CM_{max}$, CM_{max} generally denotes the strictest of the three chance model criteria, one might disregard the model selection statistics and choose M_j where $\max(\rho_j)$ if a criterion cannot do better than CM_{max} . But as model selection criteria may defy the odds by pointing to a model j where $\rho_j < \max(\rho_j)$, CM_{prop} should be used in most situations. With regard to the focus of this article, an information criterion is adequate for a certain factor level combination when the success rate is greater than the value of a given chance model criterion.

To utilise the idea of chance models, one can define an exemplary consideration set $K = \{M_1, M_2, M_3\}$ where M_1 denotes a model with $S=2$ segments (true number of segments), M_2 a model with $S=3$ segments (low overfitting) and M_3 a model with $S \geq 4$ segments (high overfitting), thus leading to the random chance criterion $CM_{ran} = 1/3 \approx 0.33$.

If a researcher has the prior probabilities $\rho_1 = 0.5$, $\rho_2 = 0.3$ and $\rho_3 = 0.2$ to observe one of

the models, the proportional chance criterion for each factor level combination is

$CM_{prop} = 0.5^2 + 0.3^2 + 0.2^2 = 0.38$, and the maximum chance criterion is

$CM_{max} = \max(0.5, 0.3, 0.2) = 0.5$.

The figures in the following section illustrate the findings of the simulation runs. Three-dimensional scatterplots are used to show the success rates of all sample/segment size combinations. The indicated values for the mixture distribution denote the relative size of the smaller segment. The line charts demonstrate the distribution of success rates for $(\pi_1^1 = 0.1, \pi_2^1 = 0.9)$, $(\pi_1^3 = 0.3, \pi_2^3 = 0.7)$ and $(\pi_1^5 = 0.5, \pi_2^5 = 0.5)$. The horizontal dotted lines illustrate the boundaries of the previously mentioned chance models with $K = \{M_1, M_2, M_3\}$: $CM_{ran} = 1/3 \approx 0.33$ (lower dotted line), $CM_{prop} = 0.38$ (medial dotted line) and $CM_{max} = 0.5$ (upper dotted line).

SUMMARY OF RESULTS

As can be seen in Figures 1 and 2, AIC only recovers the true number of segments when one of the two segments is rather small. When the smaller segment's mixture proportion values increase, success rates decrease gradually. With respect to absolute performance, AIC produces adequate solutions for $N \geq 370$ when the random chance criterion is considered. In contrast, the BIC performance (Figures 3 and 4) seems to be largely independent of the variation of the mixture proportions, showing slight advantageous performance in the presence of a niche segment. As sample size increases, success rates grow to the maximum of 100 per cent for all mixture proportions. The criterion's absolute performance is already favourable for sample sizes as low as $N=100$ against the background of the chance models.

With regard to Figures 5–8, one can see that across all mixture proportion values, CAIC successfully identifies the correct number of segments and outperforms BIC in most cases. Regardless of the existing mixture proportion, the maximum chance criterion is met for sample sizes as low as $N=100$. Compared to BIC, the

Figure 1: Sample/segment size-dependent success rates for AIC (1)

Figure 2: Sample/segment size-dependent success rates for AIC (2)

performance of CAIC is more consistent in terms of the development of the success rate across sample sizes. Whereas for BIC and CAIC, the run of the curve is concave, AIC₃ shows a convex

curve shape. For $N < 250$, this criterion shows rather low success rates that rise sharply with an increasing sample size, quickly reaching rates of almost 100 per cent. Only for AIC₃ and in the

Figure 3: Sample/segment size-dependent success rates for BIC (1)

Figure 4: Sample/segment size-dependent success rates for BIC (2)

presence of a niche segment, this surge is less abrupt. The simulation results indicate that compared to BIC and CAIC, AIC₃ performs

better in the presence of a niche segment for sample sizes of $N \geq 290$. In this range of sample sizes, AIC₃ performs very successfully.

Figure 5: Sample/segment size-dependent success rates for CAIC (1)

Figure 6: Sample/segment size-dependent success rates for CAIC (2)

Furthermore, it can be observed that for smaller sample sizes, AIC_3 meets the chance criterion standards if $\pi_1^1=0.1$ and $\pi_2^1=0.9$.

EMPIRICAL APPLICATION

To highlight the relevance of these findings, the previously evaluated criteria are applied in a

Figure 7: Sample/segment size-dependent success rates for AIC₃ (1)

Figure 8: Sample/segment size-dependent success rates for AIC₃ (2)

common marketing modelling application of mixture regression models. This example analyses influencing factors on customer satisfaction with

industrial goods. Customer satisfaction has become a fundamental and well-documented construct in marketing that is critical to business

success given its importance and established relation with customer retention and corporate profitability.^{110–112} This especially holds for capital goods in respect of industrial markets' specific characteristics.¹¹³ In a recent study, Festge and Schwaiger¹¹⁴ have developed a valid measurement scale and identified constructs that primarily drive customer satisfaction in the field of custom-designed packing machinery and systems. Using a questionnaire consisting of 15 constructs that was administered in 12 countries worldwide, the authors achieved a sample of 281 evaluations. The drivers' analysis by means of PLS path modelling revealed that only few constructs exert a significant influence on the overall customer satisfaction. The following exemplary application extends the original study by accounting for unobserved heterogeneity. In line with the simulation study, the analysis focuses on satisfaction with the following three predictors or performance features that the customers rated as the most important:

- reliability of the machines/systems;
- accuracy of the machines/systems;
- cost/performance ratio of the machines/systems.

The investigation examines a multi-attribute data analysis, where the overall satisfaction is the dependent variable and the three performance features listed above are the independent variables. All aspects were measured on 7-point rating scales with higher values denoting a higher degree of satisfaction. Using the package FlexMix,^{115,116} finite mixtures of Gaussian regression models were applied to these data for $s=1$ to $s=5$ segments. After a case-wise deletion of the missing values, the data set comprised $n=225$ observations. Table 2 presents $\ln L$ values, and

the information criteria statistics for each solution (minimum information criteria values printed in bold).

According to BIC and CAIC, a two-segment solution seems to fit the data best. In this case, 85.5 per cent of customers belong to segment 1, whereas the smaller second segment accounts for the remaining 14.5 per cent. For this sample/segment-size combination, BIC and CAIC show very high success rates (Figures 3–6) of around 95 per cent. Conversely, the success rates of AIC and AIC₃ are much lower with regard to a two-segment solution (Figures 1, 2, 7, and 8). The simulation studies by Hawkins *et al.*¹¹⁷ as well as Andrews and Currim¹¹⁸ indicate that all criteria perform considerably worse when the number of segments is increased. Furthermore, AIC has a strong tendency to overestimate the correct number of segments. In the light of these findings, the two-segment solution, as indicated by BIC and CAIC, is deemed appropriate. To illustrate the impacts of disregarding heterogeneity as well as an under and oversegmentation, segment-specific parameter estimates are calculated, including the t -values in parentheses for the one, two and three-segment solution (Table 3).

The results reveal that in an aggregate-level regression analysis of overall satisfaction in respect of the three performance features, the two variables 'reliability of the machines/systems' and 'cost/performance ratio of the machines/systems' are significantly related to overall satisfaction. Despite the stated importance, the accuracy of the machines does not exert a significant influence on satisfaction. The R^2 for the global model has a moderate value of 0.55, indicating an acceptable goodness-of-fit. A different picture emerges from the mixture regression results of a two-segment solution. The first segment exhibits a similar

Table 2: Model selection statistics

s	$\ln L$	AIC	BIC	CAIC	AIC ₃
1	–285.06	580.12	594.97	599.97	585.12
2	–260.64	543.29	575.96	586.96	554.29
3	–249.63	533.26	583.75	600.75	550.26
4	–280.33	606.67	674.98	697.98	629.67
5	–281.10	620.20	706.33	735.33	649.20

Table 3: Aggregate and segment-specific results

	Global	Mix. Reg (S=2)		Mix. Reg (S=3)		
		s=1	s=2	s=1	s=2	s=3
Intercept	0.79	0.77	1.38	1.97	2.79	0.45
Reliability of machines/systems	0.57** (10.43)	0.67** (17.23)	0.06 (0.76)	0.33** (9.37)	0.16** (5.10)	0.74** (19.52)
Accuracy of machines/systems	0.06 (1.17)	0.01 (0.21)	0.23** (2.95)	0.20** (6.38)	-0.57** (-17.99)	-0.07* (-1.75)
Cost/performance ratio of machines/systems	0.21** (4.47)	0.19** (5.94)	0.29** (3.60)	0.15** (5.20)	0.41** (14.08)	0.23** (6.89)
π_s	1	0.855	0.145	0.357	0.050	0.593

**Significant at 0.05; *significant at 0.1.

structure to the global model with no significant influence on accuracy but similar coefficients for the intercept and cost/performance ratio. The coefficient for ‘reliability of the machines/systems’, however, ranges considerably higher compared to the global model, stressing the increased importance of this feature for customers in this segment. Contrary to this, the overall satisfaction of customers in the second segment is more strongly affected by cost/performance aspects. Unlike in the other models, the accuracy of machines significantly influences the dependent variable, whereas the coefficient for ‘reliability of the machines/systems’ does not significantly differ from zero. In comparison with the global model, the R^2 in segments 1 and 2 is 0.75 and 0.58, thus demonstrating the value of the mixture regression approach.

From a managerial standpoint, these findings are more congruent with the nature of the market for industrial goods. Unlike in the consumer goods market, industrial products and services are individually produced and in some cases even engineered. These varying buyers’ needs evoke different drivers of satisfaction, which can be successfully addressed in product development and positioning decisions.¹¹⁹ For example, customers in the first segment mainly use these machines for the packing of bulk materials (eg sand) where reliability is a critical success factor. Customers in the second segment, however, use these machines for processing high-priced materials, which set higher standards for the machines’ accuracy. An aggregate-level analysis (ie undersegmentation) would not have provided

these differentiated results. Consequently, customers in the second segment might have been erroneously addressed.

Likewise, the three-segment solution provides ambiguous results. Even though the overall model fit, that is, the sum of segment-specific R^2 values weighted by relative segment size, ranges higher than in the other models ($R^2_{global} = 0.55$; $R^2_{s=2} = 0.73$; $R^2_{s=3} = 0.75$), coefficients for ‘accuracy of machines’ in $s = 2$ and $s = 3$ are counterintuitive and defy a meaningful interpretation. Furthermore, the second segment is very small and therefore not substantial enough to be addressed separately. It should be noted that such practical considerations for ruling out certain solutions need not hold for all applications and that marketing managers may well be misguided by results that appear plausible at first sight.

The results demonstrate that the interpretation of regression results in respect of the aggregate data level can be misleading, thus causing erroneous inferences. Similarly, an oversegmentation provides ambiguous results. Fitting a finite mixture model of Gaussian regressions to the data increases the model fit and provides indications for the differential targeting of each customer segment. Enriching the segmentation results with demographics enhances the interpretability of the distinctive segments and provides a useful market perspective. Most importantly for the simulation study at hand, the example shows that BIC and CAIC penalise the increase in model fit more stringently, thus indicating an appropriate solution for the analysis at hand.

DISCUSSION AND IMPLICATIONS

Owing to their considerable potential for market segmentation studies, mixture regression models have recently received increasing attention from both academics and practitioners. Besides finite mixture conjoint models, the FIMIX-PLS approach to response-based clustering in variance-based path modelling constitutes one of the most promising application areas of mixture regression models.^{120,121} The price paid for the flexibility in the application of these models is that the inference is often a challenge.¹²² One fundamental difficulty in the application is related to the problem of model selection, that is, the choice of the number of segments. Retaining the correct number of segments is, however, crucial because many managerial decisions rely on this.¹²³ From a managerial standpoint, if consumers are truly heterogeneous, for example, with regard to their responses to promotional activities, an under- or oversegmentation might lead to erroneous estimations of the response to marketing efforts, which will affect sales forecasts incorrectly. Given that companies spend millions of dollars on market targeting and positioning, these problems are nontrivial. Consequently, it is important to adequately capture this kind of heterogeneity in marketing models.

Since the correct number of segments is unknown in real-world applications, a thorough understanding of measures that guide the model selection decision is of fundamental importance. Researchers who take the wrong measure into consideration when deriving a market segmentation strategy may be misguided. This paper addresses this problem by evaluating how the interaction of the most important influencing factors for the measures' success — sample and segment size — affect the performance of four of the most widely used information criteria for assessing the correct number of segments in mixture regression models. For the first time, the quality of these criteria is evaluated for a wide spectrum of possible constellations. Furthermore, relative and absolute performances are analysed with regard to outside criteria. The results induce recommendations on criterion selection when a certain sample size is given and help to judge

what sample size is needed in order to guarantee an accurate decision based on a certain criterion. Furthermore, the results also demonstrate that in the presence of certain sample/segment size combinations, decisions grounded on a specific criterion might prove problematic.

AIC presents an extremely poor performance across all simulation situations. From an application-oriented point of view and taking into account the high percentage of studies relying on this criterion to assess the number of segments in the model, this poor performance proves to be problematic. More precisely, the poor performance makes the appropriateness of these studies' results highly questionable. With regard to AIC, the results are contrary to the findings by Oliveira-Brochado and Martins,¹²⁴ who maintain that this criterion performs well in a simulation design with equal segment sizes. In addition, AIC performs much worse in this study than in Andrews and Currim.¹²⁵ CAIC performs favourably, showing slight weaknesses in determining the true number of segments for high sample sizes in the presence of a niche segment. In the latter situation, AIC₃ performs well, quickly achieving success rates of over 90 per cent, hence meeting random chance, proportional chance and maximum chance boundaries. In contrast to previous findings by Andrews and Currim,¹²⁶ CAIC outperforms BIC across all sample/segment size combinations, whereupon the deviation is marginal when the segments of the mixture are not well separated ($\pi_1^1 = 0.1$ and $\pi_1^2 = 0.2$ and). AIC₃ shows a relative and absolute positive performance for high levels of sample size, which previous studies have not considered. Interestingly, most criteria perform better in the presence of a niche segment, which is an unexpected finding, since one expects the existence of a small segment to add complexity to the retention problem. This finding could be attributed to the design of the second (niche) segment with regard to the sizable separation between the regression coefficients of both segments.

However, as the results do not reflect the efficacy of the criteria across all possible model constellations, it is important to interpret these

results with caution. The scope of this study was to evaluate the interaction of sample and segment size across a broad range of levels. Consequently, several other influencing factors were held constant, which might have biased the results. For example, the simulation design did not control for correlations between the three predictor variables, nor was the fuzziness, that is, the degree of separation between the segments, investigated.

Consequently, subsequent simulation studies should systematically vary these influencing factors to assess their influence on the criteria's performance. Continued research is needed on the performance of model selection criteria to provide practical guidelines for ascertaining the correct number of segments in a mixture and to guarantee accurate conclusions for marketing practice.

However, considering the great number of research projects, one should be critical of the idea of finding a unique measure that can be considered optimal in every simulation design or even practical applications, as other studies have already indicated. Model selection decisions should rather be based on various evidences, not only derived from the data at hand but also from practical considerations. Well-known considerations, by means of which the effectiveness of segmentation results can be evaluated, include that segments are profitable enough and can be effectively reached. Furthermore, they need to be actionable.¹²⁷ Researchers, however, also face situations in which practical considerations provide ambiguous or contradictory results. In such situations, the consideration of an appropriate model selection criterion can serve as a focal point for management decisions.

Additional elaborations of the possibility to include *a priori* information or the expected costs of under or oversegmentation directly into the design of model selection criteria are required to foster application and adoption by practitioners. The ultimate goal of such efforts is to merge data and theory-driven assessment of marketing problems. The integration of *a priori* information might enhance the plausibility of the results and support the diffusion of this very promising technique in marketing practice.

Acknowledgments

The author would like to thank the anonymous reviewers for their helpful comments.

References

- 1 Wedel, M. and Kamakura, W. A. (2000) 'Market Segmentation. Conceptual and Methodological Foundations', 2nd edn, International Series in Quantitative Marketing, Kluwer Academic Publishers, Boston, Dordrecht, London.
- 2 *Ibid.*
- 3 McLachlan, G. J. and Peel, D. (2000) 'Finite Mixture Models', Wiley-Interscience, New York, NY.
- 4 Oh, M. S. and Raftery, A. (2003) 'Model-based clustering with dissimilarities: A Bayesian approach', Technical Report, No. 441, Department of Statistics, University of Washington.
- 5 DeSarbo, W. S., Jedidi, K. and Sinha, I. (2001) 'Customer value analysis in a heterogeneous market', *Strategic Management Journal*, Vol. 22, No. 9, pp. 845–857.
- 6 Wedel, M. and DeSarbo, W. S. (2002) 'Market segment derivation and profiling via a finite mixture model framework', *Marketing Letters*, Vol. 13, No. 1, pp. 17–25.
- 7 Reinartz, W., Thomas, J. S. and Kumar, V. (2005) 'Balancing acquisition and retention resources to maximize customer profitability', *Journal of Marketing*, Vol. 69, No. 1, pp. 63–79.
- 8 Srinivasan, R. (2006) 'Dual distribution and intangible firm value: Franchising in restaurant chains', *Journal of Marketing*, Vol. 70, No. 3, pp. 120–135.
- 9 Andrews, R. L., Currim, I. S., Leeflang, P. and Lim, J. (2007) 'Estimating the SCAN*PRO model of store sales: HB, FM or just OLS?', *International Journal of Research in Marketing*, Vol. 25, No. 1, pp. 22–33.
- 10 Hahn, C. H., Johnson, M. D., Herrmann, A. and Huber, F. (2002) 'Capturing customer heterogeneity using a finite mixture PLS approach', *Schmalenbach Business Review*, Vol. 54, No. 3, pp. 243–269.
- 11 Ringle, C. M., Wende, S. and Will, A. (2005) 'Customer segmentation with FIMIX-PLS', in Aluja, T., Casanovas, J., Esposito Vinzi, V., Morrineau, A. and Tenenhaus, M. (eds), 'PLS and Related Methods – Proceedings of the PLS'05 International Symposium', Decisia, Paris, pp. 507–514.
- 12 Ringle, C. M., Wende, S. and Will, A. (2008) 'The finite mixture partial least squares approach: Methodology and application', in Esposito Vinzi, V., Chin, W.W., Henseler, J. and Wang, H. (eds), 'Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields', Springer, Berlin, forthcoming.
- 13 Hahn, C. H., Johnson, M. D., Herrmann, A. and Huber, F. (2002) *op. cit.*
- 14 Ringle, C. M., Sarstedt, M. and Mooi, E. A. (2008) 'Response-based segmentation using finite mixture partial least squares. Theoretical foundations applied to measuring customer satisfaction', *Annals of Information Systems*, forthcoming.
- 15 Ringle, C. M. (2006) 'Segmentation for path models and unobserved heterogeneity: The finite mixture partial least squares approach', *Research Papers on Marketing and Retailing*, No. 35, available at: <http://www.ibl-uni-hh.de/RP035.pdf> (accessed 28th May, 2008).
- 16 Ringle, C. M., Wende, S. and Will, A. (2008) *op. cit.*

- 17 Bouncken, R. and Koch, M. (2006) 'Inter-organisationales Vertrauen und Ergebnisse von Kooperationen: Eine empirische Untersuchung mittels Finite-Mixture-PLS', in Bauer, H.M., Neumann, M.M. and Schüle, A. (eds), 'Konsumentenvertrauen, Konzepte und Anwendungen für ein nachhaltiges Kundenbindungsmanagement', Verlag Vahlen, München, pp. 265–277.
- 18 Grasmugg, S. (2006) 'Mass Customization als strategische Anwendung des Electronic Business. Eine empirische Untersuchung zu Status, Determinanten und Erfolgswirksamkeit', Josef Eul Verlag, Lohmar.
- 19 Sarstedt, M. (2008) 'Treating unobserved heterogeneity in PLS path modeling. A comparison of FIMIX-PLS with different data analysis strategies', in 'Proceedings of the 2008 Global Marketing Conference of the Korean Academy of Marketing Science (KAMS)', Shanghai, China.
- 20 Sarstedt, M., Ringle, C. M., Schloderer, M. P. and Schwaiger, M. (2008) 'Accounting for unobserved heterogeneity in the analysis of antecedents and consequences of corporate reputation: An application of FIMIX-PLS', in 'Proceedings of the 37th Annual Conference of the European Marketing Academy (EMAC)', Brighton, Great Britain.
- 21 Andrews, R. L., Ansari, A. and Currim, I. S. (2002) 'Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction and pathworth recovery', *Journal of Marketing Research*, Vol. 39, No. 1, pp. 87–98.
- 22 Boone, D. S. and Roehm, M. (2002) 'Evaluating the appropriateness of market segmentation solutions using artificial neural networks and the membership clustering criterion', *Marketing Letters*, Vol. 13, No. 4, pp. 317–333.
- 23 Andrews, R. L. and Currim, I. S. (2003a) 'Retention of latent segments in regression-based marketing models', *International Journal of Research in Marketing*, Vol. 20, No. 4, pp. 315–321.
- 24 Soromenho, G. (1994) 'Comparing approaches for testing the number of components in a finite mixture model', *Computational Statistics*, Vol. 9, No. 1, pp. 65–78.
- 25 Nylund, K. L., Asparouhov, T. and Muthén, B. O. (2006) 'Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study', White Paper, available at: http://www.statmodel.com/download/LCA_tech11_nylund_v83.pdf (accessed 28th May, 2008).
- 26 Henson, J. M., Reise, S. P. and 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*, Vol. 14, No. 2, pp. 202–226.
- 27 Tofighi, D. and Enders, C. K. (2007) 'Identifying the correct number of classes in growth mixture models', in Hancock, G. R. and Samuelson, K.M. (eds), 'Advances in Latent Variable Mixture Models', Information Age Publishing, Greenwhich, pp. 317–341.
- 28 Hawkins, D. S., Allen, D. M. and Stromberg, A. J. (2001) 'Determining the number of components in mixtures of linear models', *Computational Statistics & Data Analysis*, Vol. 38, No. 1, pp. 15–48.
- 29 Andrews, R. L. and Currim, I. S. (2003b) 'A comparison of segment retention criteria for finite mixture logit models', *Journal of Marketing Research*, Vol. 40, No. 2, pp. 235–243.
- 30 Oliveira-Brochado, A. and Martins, F.V. 'Examining the segment retention problem for the "group satellite" case', FEP Working Papers, No. 220, available at: http://www.fep.up.pt/investigacao/workingpapers/06.07.04_WP220_brochadomartins.pdf (accessed 28th May, 2008).
- 31 Wedel, M. and Kamakura, W. A. (2000) *op. cit.*
- 32 Aitkin, M. and Rubin, D. B. (1985) 'Estimation and hypothesis testing in finite mixture models', *Journal of the Royal Statistical Society, Series B, Methodological*, Vol. 47, No. 1, pp. 67–75.
- 33 Everitt, B. S. (1981) 'A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distributions', *Multivariate Behavioral Research*, Vol. 16, No. 2, pp. 171–180.
- 34 Everitt, B. S. (1988) 'A Monte Carlo investigation of the likelihood ratio test for number of classes in latent class analysis', *Multivariate Behavioral Research*, Vol. 23, No. 4, pp. 531–538.
- 35 McLachlan, G. J. and Peel, D. (2000) *op. cit.*
- 36 Nylund, K. L., Asparouhov, T. and Muthén, B. O. (2006) *op. cit.*
- 37 McLachlan, G. J. (1987) 'On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture', *Applied Statistics*, Vol. 36, No. 3, pp. 318–324.
- 38 Wedel, M. and Kamakura, W. A. (2000) *op. cit.*
- 39 Lo, Y., Mendell, N. R. and Rubin, D. B. (2001) 'Testing the number of components in a normal mixture', *Biometrika*, Vol. 88, No. 3, pp. 767–778.
- 40 Jeffries, N. O. (2003) 'A note on testing the number of components in a normal mixture', *Biometrika*, Vol. 90, No. 4, pp. 991–994.
- 41 Oliveira-Brochado, A. and Martins, F.V. (2006) *op. cit.*
- 42 Akaike, H. (1973) 'Information theory and an extension of the maximum likelihood principle', in Petrov, B.N. and Csaki, F. (eds), 'Budapest: Second International Symposium on Information Theory', Akademiai Kiadó, Budapest, pp. 267–281.
- 43 Bozdogan, H. (1987) 'Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions', *Psychometrika*, Vol. 52, No. 3, pp. 345–370.
- 44 Schwarz, G. (1978) 'Estimating the dimensions of a model', *The Annals of Statistics*, Vol. 6, No. 2, pp. 461–464.
- 45 Harzing, A. -W. (2006) 'Journal quality list', White Paper, 23rd edn, available at: <http://www.harzing.com/download/jql.zip> (accessed November 20, 2006).
- 46 Bell, D. R. and Lattin, J. M. (2000) 'Looking for loss aversion in scanner panel data: The confounding effect of price response heterogeneity', *Marketing Science*, Vol. 19, No. 2, pp. 185–200.
- 47 Heilman, C. M., Bowman, D. and Wright, G. P. (2000) 'The evolution of brand preferences and choice behaviors of consumers new to a market', *Journal of Marketing Research*, Vol. 37, No. 2, pp. 139–155.
- 48 Mazumdar, T. and Papatla, P. (2000) 'An investigation of reference prices segments', *Journal of Marketing Research*, Vol. 37, No. 2, pp. 246–258.
- 49 Shachar, R. and Emerson, J. W. (2000) 'Cast demographics, unobserved segments, and heterogeneous switching costs in a television viewing choice model', *Journal of Marketing Research*, Vol. 37, No. 2, pp. 173–186.
- 50 Danaher, P. J. and Mawhinney, D. F. (2001) 'Optimizing television program schedules using choice modeling', *Journal of Marketing Research*, Vol. 38, No. 3, pp. 298–312.
- 51 Erdem, T., Mayhew, G. and Sun, B. (2001) 'Understanding reference-price shoppers: A within- and cross-category analysis', *Journal of Marketing Research*, Vol. 38, No. 4, pp. 445–457.
- 52 Gönül, F. F., Carter, E., Petrova, E. and Srinivasan, K. (2001) 'Promotion of prescription drugs and its impact on physicians' choice behavior', *Journal of Marketing*, Vol. 65, No. 3, pp. 79–90.

- 53 Thomas, J. S. (2001) 'A methodology for linking customer acquisition to customer retention', *Journal of Marketing Research*, Vol. 38, No. 2, pp. 262–268.
- 54 Andrews, R. L., Ainsle, A. and Currim, I. S. (2002) 'An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity', *Journal of Marketing Research*, Vol. 39, No. 4, pp. 479–487.
- 55 Andrews, R. L., Ansari, A. and Currim, I. S. (2002) *op. cit.*
- 56 Danaher, P. J. (2002) 'Optimal pricing of new subscription services: Analysis of a market experiment', *Marketing Science*, Vol. 21, No. 2, pp. 119–138.
- 57 Hofstede, F. T., Kim, Y. and Wedel, M. (2002) 'Bayesian prediction in hybrid conjoint analysis', *Journal of Marketing Research*, Vol. 39, No. 2, pp. 253–261.
- 58 Hofstede, F. T., Wedel, M. and Steenkamp, J. -B. E. M. (2002) 'Identifying spatial segments in international markets', *Marketing Science*, Vol. 21, No. 2, pp. 160–177.
- 59 Papadla, P. and Bhatnagar, A. (2002) 'Shopping style segmentation of consumers', *Marketing Letters*, Vol. 13, No. 2, pp. 91–106.
- 60 Wedel, M. and DeSarbo, W. S. (2002) *op. cit.*
- 61 Agarwal, M. K. (2003) 'Developing global segments and forecasting market shares, a simultaneous approach using survey data', *Journal of International Marketing*, Vol. 11, No. 4, pp. 56–80.
- 62 Chung, J. and Rao, V. R. (2003) 'A general choice model for bundles with multiple-category products: Application to market segmentation and optimal pricing for bundles', *Journal of Marketing Research*, Vol. 40, No. 2, pp. 115–130.
- 63 Danaher, P. J., Wilson, I. P. and Davis, R. A. (2003) 'A comparison of online and offline consumer brand loyalty', *Marketing Science*, Vol. 22, No. 4, pp. 461–476.
- 64 Ho, T. -H. and Chong, J. -K. (2003) 'A parsimonious model of stockkeeping-unit choice', *Journal of Marketing Research*, Vol. 40, No. 3, pp. 351–365.
- 65 Wu, J. and Rangaswamy, A. (2003) 'A fuzzy set model of search and consideration with an application to an online market', *Marketing Science*, Vol. 22, No. 3, pp. 411–434.
- 66 Anand, B. N. and Shachar, R. (2004) 'Brands as beacons: A new source of loyalty to multiproduct firms', *Journal of Marketing Research*, Vol. 41, No. 2, pp. 135–150.
- 67 Bowman, D., Heilman, C. M. and Seetharaman, P. B. (2004) 'Determinants of product-use compliance behavior', *Journal of Marketing Research*, Vol. 41, No. 3, pp. 324–338.
- 68 Lewis, M. (2004) 'The influence of loyalty programs and short-term promotions on customer retention', *Journal of Marketing Research*, Vol. 41, No. 3, pp. 281–292.
- 69 Varki, S. and Chintagunta, P. K. (2004) 'The augmented latent class model: Incorporating additional heterogeneity in the latent class model for panel data', *Journal of Marketing Research*, Vol. 41, No. 2, pp. 226–233.
- 70 Zhang, J. and Krishnamurthi, L. (2004) 'Customizing promotions in online stores', *Marketing Science*, Vol. 23, No. 4, pp. 561–578.
- 71 Jedidi, K. and Kohli, R. (2005) 'Probabilistic subset-conjunctive models for heterogeneous consumers', *Journal of Marketing Research*, Vol. 42, No. 4, pp. 483–494.
- 72 Lewis, M. (2005) 'Incorporating strategic consumer behavior into customer valuation', *Journal of Marketing*, Vol. 69, No. 4, pp. 230–238.
- 73 Reinartz, W., Thomas, J. S. and Kumar, V. (2005) *op. cit.*
- 74 Rust, R. T. and Verhoef, P. C. (2005) 'Optimizing the marketing interventions mix in intermediate-term C RM', *Marketing Science*, Vol. 24, No. 3, pp. 477–489.
- 75 Thomas, J. S. and Sullivan, U. Y. (2005) 'Managing marketing communications with multichannel customers', *Journal of Marketing*, Vol. 69, No. 4, pp. 239–251.
- 76 Kivetz, R., Urminsky, O. and Zheng, Y. (2006) 'The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention', *Journal of Marketing Research*, Vol. 43, No. 1, pp. 39–58.
- 77 Mantrala, M. K., Seetharaman, P. B., Kaul, R., Gopalakrishna, S. and Stam, A. (2006) 'Optimal pricing strategies for an automotive aftermarket retailer', *Journal of Marketing Research*, Vol. 43, No. 4, pp. 588–604.
- 78 Srinivasan, R. (2006) *op. cit.*
- 79 Ho, T. -H. and Chong, J. -K. (2003) *op. cit.*
- 80 Danaher, P. J. and Mawhinney, D. F. (2001) *op. cit.*
- 81 Danaher, P. J. (2002) *op. cit.*
- 82 Wu, J. and Rangaswamy, A. (2003) *op. cit.*
- 83 Bucklin, R. E. and Gupta, S. (1992) 'Brand choice, purchase incidence, and segmentation: An integrated modeling approach', *Journal of Marketing Research*, Vol. 29, No. 2, pp. 201–215.
- 84 Andrews, R. L., Ainsle, A. and Currim, I. S. (2002) *op. cit.*
- 85 Andrews, R. L. and Currim, I. S. (2003a) *op. cit.*
- 86 Bozdogan, H. (1994) 'Mixture-model cluster analysis using model selection criteria and a new information measure of complexity', in Bozdogan, H. (ed.), 'Proceedings of the First US/Japan Conference on Frontiers of Statistical Modelling: An Informational Approach', Vol. 2, Kluwer Academic Publishers, Boston, Dordrecht, London, pp. 69–113.
- 87 Vermunt, J. K. and Magidson, J. (2005) 'Latent gold 4.0 – User's guide', available at: <http://www.statisticalinnovations.com/products/fullmanual.pdf> (accessed 28th May, 2008).
- 88 Muthén, L. K. and Muthén, B. O. (2006) 'Mplus. Statistical Analysis with Latent Variables – User's Guide Version 4', Muthén & Muthén, Los Angeles, CA.
- 89 Leisch, F. (2004) 'FlexMix: A general framework for finite mixture models and latent class regression in R', *Journal of Statistical Software*, Vol. 11, No. 8, pp. 1–18.
- 90 Grün, B. and Leisch, F. (2007) 'Fitting finite mixtures of linear regressions in R', *Computational Statistics & Data Analysis*, Vol. 51, No. 11, pp. 5247–5252.
- 91 Andrews, R. L. and Currim, I. S. (2003a) *op. cit.*
- 92 Bozdogan, H. (1990) 'On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models', *Communications in Statistics, Theory and Methods*, Vol. 19, No. 1, pp. 221–278.
- 93 Bozdogan, H. (1993) 'Choosing the number of component clusters in the mixture-model using a new information complexity criterion of the inverse-fisher information matrix', in Opitz, O. and Klar, R. (eds), 'Information and Classification', Springer, Heidelberg, pp. 40–54.
- 94 Celeux, G. and Soromenho, G. (1996) 'An entropy criterion for assessing the number of clusters in a mixture model', *Journal of Classification*, Vol. 13, No. 2, pp. 195–212.
- 95 Hawkins, D. S., Allen, D. M. and Stromberg, A. J. (2001) *op. cit.*
- 96 Andrews, R. L. and Currim, I. S. (2003a) *op. cit.*
- 97 Oliveira-Brochado, A. and Martins, F. V. (2006) *op. cit.*
- 98 Andrews, R. L. and Currim, I. S. (2003a) *op. cit.*
- 99 Hawkins, D. S., Allen, D. M. and Stromberg, A. J. (2001) *op. cit.*
- 100 Oliveira-Brochado, A. and Martins, F. V. (2006) *op. cit.*
- 101 Hawkins, D. S., Allen, D. M. and Stromberg, A. J. (2001) *op. cit.*
- 102 Leisch, F. (2004) *op. cit.*
- 103 Grün, B. and Leisch, F. (2007) *op. cit.*

- 104 Dempster, A. P., Laird, A. M. and Rubin, D. B. (1977) 'Maximum likelihood from incomplete data via the EM-algorithm', *Journal of the Royal Statistical Society, Series B*, Vol. 39, No. 1, pp. 1–38.
- 105 R Development Core Team (2007) 'R: A language and environment for statistical computing. R foundation for statistical computing', available at: <http://www.R-project.org> (accessed 28th May, 2008).
- 106 Grün, B. and Leisch, F. (2008) 'Applications of finite mixtures of regression models', White Paper, R Development Core Team, available at: <http://cran.r-project.org/web/packages/flexmix/vignettes/regression-examples.pdf> (accessed 28th May, 2008).
- 107 Wedel, M. and Kamakura, W. A. (2000) *op. cit.*
- 108 Morrison, D. G. (1969) 'On the interpretation of discriminant analysis', *Journal of Marketing Research*, Vol. 6, No. 2, pp. 156–163.
- 109 *Ibid.*
- 110 Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J. and Bryant, B. E. (1996) 'The American customer satisfaction index: Nature, purpose, and findings', *Journal of Marketing*, Vol. 60, No. 4, pp. 7–18.
- 111 Mittal, V., Anderson, E. W., Sayrak, A. and Tadikamalla, P. (2005) 'Dual emphasis and the long-term financial impact of customer satisfaction', *Marketing Science*, Vol. 24, No. 5, pp. 544–555.
- 112 Morgan, N. A., Anderson, E. W. and Mital, V. (2005) 'Understanding firms' customer satisfaction information usage', *Journal of Marketing*, Vol. 69, No. 3, pp. 131–151.
- 113 Homburg, C. and Rudolph, B. (2001) 'Customer satisfaction in industrial markets: Dimensional and multiple role issues', *Journal of Business Research*, Vol. 52, No. 1, pp. 15–33.
- 114 Festge, F. and Schwaiger, M. (2007) 'The drivers of customer satisfaction with industrial goods: An international study', *Advances in International Marketing*, Vol. 18, pp. 183–212.
- 115 Leisch, F. (2004) *op. cit.*
- 116 Grün, B. and Leisch, F. (2007) *op. cit.*
- 117 Hawkins, D. S., Allen, D. M. and Stromberg, A. J. (2001) *op. cit.*
- 118 Andrews, R. L. and Currim, I. S. (2003a) *op. cit.*
- 119 Festge, F. and Schwaiger, M. (2007) *op. cit.*
- 120 Hahn, C. H., Johnson, M. D., Herrmann, A. and Huber, F. (2002) *op. cit.*
- 121 Ringle, C. M., Wende, S. and Will, A. (2008) *op. cit.*
- 122 Frühwirth-Schnatter, S. (2006) 'Finite Mixture and Markov Switching Models', Springer, New York.
- 123 Andrews, R. L. and Currim, I. S. (2003a) *op. cit.*
- 124 Oliveira-Brochado, A. and Martins, F. V. (2006) *op. cit.*
- 125 Andrews, R. L. and Currim, I. S. (2003a) *op. cit.*
- 126 *Ibid.*
- 127 Kotler, P. and Keller, K. L. (2006) 'Marketing Management', 12th edn, Pearson Prentice Hall, Upper Saddle River.