

## Repositório ISCTE-IUL

---

Deposited in *Repositório ISCTE-IUL*:

2023-03-07

Deposited version:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Fonseca, J. R. S. & Cardoso, M. G. M. S. (2005). Retail clients latent segments. In Bento, C., Cardoso, A., and Dias, G. (Ed.), *Progress in Artificial Intelligence. EPIA 2005. Lecture Notes in Computer Science*. (pp. 348-358). Covilhã, Portugal: Springer.

Further information on publisher's website:

[10.1007/11595014\\_35](https://doi.org/10.1007/11595014_35)

Publisher's copyright statement:

This is the peer reviewed version of the following article: Fonseca, J. R. S. & Cardoso, M. G. M. S. (2005). Retail clients latent segments. In Bento, C., Cardoso, A., and Dias, G. (Ed.), *Progress in Artificial Intelligence. EPIA 2005. Lecture Notes in Computer Science*. (pp. 348-358). Covilhã, Portugal: Springer., which has been published in final form at [https://dx.doi.org/10.1007/11595014\\_35](https://dx.doi.org/10.1007/11595014_35). This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

---

### Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

---

# Retail Clients Latent Segments

Jaime R. S. Fonseca<sup>1</sup>, and Margarida G. M. S. Cardoso<sup>2</sup>

<sup>1</sup>ISCSP-Instituto Superior de Ciências Sociais e Políticas,  
R. Almerindo Lessa, Pólo Universitário do Alto da Ajuda,  
1349-055 Lisboa. Portugal  
[jaimefonseca@iscsp.utl.pt](mailto:jaimefonseca@iscsp.utl.pt)

<sup>2</sup>ISCTE – Business School. Department of Quantitative Methods.  
Av. das Forças Armadas 1649-026 Lisboa. Portugal  
[margarida.cardoso@iscte.pt](mailto:margarida.cardoso@iscte.pt)

**Abstract.** Latent Segments Models (LSM) are commonly used as an approach for market segmentation. When using LSM, several criteria are available to determine the number of segments. However, it is not established which criteria are more adequate when dealing with a specific application. Since most market segmentation problems involve the simultaneous use of categorical and continuous base variables, it is particularly useful to select the *best* criteria when dealing with LSM with mixed type base variables. We first present an empirical test, which provides the ranking of several information criteria for model selection based on ten mixed data sets. As a result, the ICL-BIC, BIC, CAIC and  $\mathcal{L}$  criteria are selected as the *best* performing criteria in the estimation of mixed mixture models. We then present an application concerning a retail chain clients' segmentation. The *best* information criteria yield two segments: *Preferential Clients* and *Occasional Clients*. Keywords: Clustering, Finite Mixture Models, Information Criteria, Marketing Research

## Introduction

Market segmentation is the division of a heterogeneous market into homogeneous sub-markets of consumers, clusters or segments, with similar behaviour within segments and different behaviour across segments. The first application of market segmentation emerged in 1956 [38].

Segmentation is an essential instrument of marketing [28], [39]. It provides a better market understanding and, consequently, means to develop more successful business strategies.

A market segmentation solution is a function of the market segmentation variables and of a specific segmentation (clustering) procedure. In what concerns base variables for segmentation, product-specific variables (see table 1) should be considered [39]. Other attributes may help profiling the segments' structures (for an overview of previous works on the use of demographics, psychographics, and other variables in segmentation studies, see [21], [28], [42]).

**Table 1.** Segmentation base variables (some examples)

General observable variables	Product-specific observable variables	General unobservable variables	Product-specific unobservable variables
Demographic	Usage frequency	Life-style	Benefits
Socio-economic	Brand loyalty	Psychographics	Utilities
Cultural	User status	Personality	Preferences
Geographic	Situations	Values	Intentions
			Perceptions
			Attributes

Along with the selection of a set of potential segmentation variables, a segmentation procedure must be chosen, which delivers a segmentation solution.

In this paper we present the segmentation of clients of a retail chain which is based on product-specific variables and results from the estimation of a Latent Segments Model (LSM) [14], [42]. This approach enables the simultaneous use of categorical and continuous segmentation base variables. It is a probabilistic clustering approach which assumes that the variables' observations in a sample arise from different segments of unknown proportions. Estimation of the LSM is typically based on maximum likelihood.

## Latent Segments Models

The aim of Latent Segments Models [14] (or Finite Mixture Models) is to identify the latent segments required to explain the associations among a set of observed variables (segmentation base variables) and to allocate observations to these segments.

The use of LSM has become increasingly popular in the marketing literature [18], [42]. This approach to segmentation offers some advantages when compared with other techniques: it identifies market segments[19]; it provides means to select the number of segments [30]; it is able to deal with diverse types of data (different measurement levels) [40]; it outperforms more traditional approaches [41].

LSM provide a clustering model based approach, a statistical model being postulated for the population from which the sample under study is coming, and assuming that the data is generated by a mixture of underlying (density) probability distributions.

Let  $\underline{y}_i = (y_{ip})$  be the vector representing the scores of the  $i$ th case for the  $p$ th segmentation base variable ( $i = 1, \dots, n$ ;  $p = 1, \dots, P$ ).

Several types of segmentation variables may be considered which have a conditional (within-cluster) distribution in the exponential family (such as Bernoulli, Poisson, Multinomial or Normal distribution) [24], [25], [32], [40]. Considering  $S$  as the number of segments and  $s=1, \dots, S$ , we define  $\lambda_s$  as mixing probability or segment size. Assuming local independence [40], the LSM may generally be presented as follows:

$$f(\underline{y}_i | \underline{\psi}) = \sum_{s=1}^S \lambda_s \prod_{p=1}^P f_s(y_{ip} | \underline{\theta}_{sp}) \quad (1)$$

where  $i = 1, \dots, n$ ,

$$\sum_{s=1}^S \lambda_s = 1, \quad \underline{\psi} = \{\underline{\lambda}, \Theta\}, \quad \text{with } \underline{\lambda} = \{\lambda_1, \dots, \lambda_{S-1}\} \text{ and } \Theta = \{\underline{\theta}_{1p}, \dots, \underline{\theta}_{sp}\}.$$

When considering mixed type variables for segmentation we may additionally specify that

$$f_s(y_{ip} | \underline{\theta}_{sp}) \sim N(\mu_{sp}, \sigma_{sp}^2)$$

for each one of the continuous attributes, and

$$f_s(y_{ip} | \underline{\theta}_{sp}) \sim \text{Mult}_{C_p}(1; \theta_{sp1}, \dots, \theta_{spC_p}),$$

for the categorical attributes, with  $C_p$  categories (*e.g.* see [25]).

Although continuous attributes could be categorized and also modelled by the multinomial distribution, this may result in considerable loss of information [13]. Furthermore it is difficult to establish an adequate number of categories [12]; however, discretization may be very useful in particular when continuous variables which do not belong to the exponential family are considered.

The LSM assumption of conditional independence can be relaxed by using the appropriate multivariate rather than univariate distributions for sets of locally dependent variables: multivariate normal distribution for sets of continuous variables and a set of categorical variables can be combined into a joint multinomial distribution.

The LSM estimation problem simultaneously addresses the estimation of distributional parameters and classification of cases into segments, yielding mixing probabilities.

Finally, modal allocation provides means for constituting a partition assigning each case to the segment with the highest posterior probability which is given by  $\text{Max}_{s=1, \dots, S} \tau_{is}$ , where

$$\tau_{is} = \hat{\lambda}_s(\underline{y}_i | \underline{\psi}) = \frac{\hat{\lambda}_s \prod_{p=1}^P f_s(\underline{y}_{ip} | \hat{\theta}_{sp})}{\sum_{s=1}^S \hat{\lambda}_s \prod_{p=1}^P f_s(\underline{y}_{ip} | \hat{\theta}_{sp})} \quad (2)$$

Maximum likelihood estimates of the vector parameter  $\underline{\psi}$  can be obtained by treating the unobserved segment labels as missing data and using the EM algorithm [22], [30], [34].

## Model Selection Criteria

### 3.1 Introduction

Several criteria may be considered for the selection of Latent Segments Models (LSM). In the present work we consider theoretical information based criteria. These criteria are generally based on the likelihood function (which we want to maximize) and a measure of model complexity (which we want to minimize). Thus, all theoretical information criteria balance parsimony (fitting a model with a large number of components requires the estimation of a very large number of parameters and a potential loss of precision in these estimates [29]), and model complexity (which tends to improve the model fit to the data).

The general form of information criteria is as follows:

$$-2 \log L(\hat{\psi}) + C \quad (3)$$

where the first term measures the lack of fit and the second term  $C$  includes a measure for model complexity, and a penalization factor. Some information criteria are shown on table 2.

The emphasis on information criteria begins with the pioneer work of Akaike [2]. Akaike's Information Criterion (AIC) chooses a model with  $S$  segments that minimises (3) with  $C = 2 n_{\psi}$ .

Later, Bozdogan [8] suggested the modified AIC criterion (AIC3) in the context of mixture models, using 3 instead of 2 on penalizing term; so, it chooses a model with  $S$  segments that minimises (3) with  $C = 3 n_{\psi}$ .

Another variant of AIC, the corrected AIC, is proposed [26], focusing on the small-sample bias adjustment (AIC may perform poorly if there are too many parameters in relation to the sample size); AICc thus selects a model with  $S$  segments that minimises (3) with  $C = 2n_{\psi} \left( \frac{n}{n-n_{\psi}-1} \right)$ .

A new criterion is then proposed - AICu - because AICc still tends to over fit as the sample size increase [31].

With the consistent AIC criterion (CAIC), Bozdogan [9] noted that the term  $n_{\psi}(1+\log n)$  has the effect of increasing the penalty term and, as a result, minimization of CAIC leads in general to models with fewer parameters than AIC does.

The Bayesian information criterion (BIC) was proposed by Schwarz [36], and chooses a model with  $S$  segments that minimises (3) with  $C = n_{\psi} \log n$ ; in a different way, from the notion of stochastic complexity, Rissanen [35] proposed an equivalent criterion in form, the minimum descriptive length (MDL).

**Table 2.** Some theoretical information criteria

Criterion	Definition	Reference
AIC	$-2LL + 2n_{\underline{\psi}}$	[1]
AIC3	$-2LL + 3n_{\underline{\psi}}$	[8]
AICc	$AIC + (2n_{\underline{\psi}}(n_{\underline{\psi}} + 1))/(n - n_{\underline{\psi}} - 1)$	[26]
AICu	$AICc + n \log(n/(n - n_{\underline{\psi}} - 1))$	[31]
CAIC	$-2LL + n_{\underline{\psi}}(1 + \log n)$	[9]
BIC/MDL	$-2LL + n_{\underline{\psi}} \log n$	[36] / [35]
CLC	$-2LL + 2EN(S)$	[4]
ICL-BIC	$BIC + 2EN(S)$	[5]
NEC	$NEC(S) = EN(S)/(L(S) - L(1))$	[6]
AWE	$-2LL_c + 2n_{\underline{\psi}}(3/2 + \log n)$	[3]
$\mathcal{L}$	$-LL + (n_{\underline{\psi}}/2) \sum \log(n\lambda_s/12) + S/2 \log(n/12) + S(n_{\underline{\psi}} + 1)/2$	[22]
S	Number of segments	
n	Number of observations	
$n_{\underline{\psi}}$	Number of model parameters	
L	Likelihood function	
LL	Log Likelihood function	
LL <sub>c</sub>	Classification Log Likelihood	
$EN(S)$	Entropy measure	

The CLC - Complete Likelihood Classification criterion [30] is proposed as an approximation of the classification likelihood criterion [7]. It chooses a model with S segments that minimises (3) with  $C = 2EN(S)$ , where the term  $2EN(S)$  penalizes poorly separated segments, with

$$EN(S) = - \sum_{i=1}^n \sum_{s=1}^S \tau_{is} \log \tau_{is}$$

A particular approximation version of the integrated classification likelihood criterion (ICL) referred to as ICL-BIC by McLachlan and Peel [30], chooses a model with S segments that minimises (3) with  $C = 2EN(S) + n_{\underline{\psi}} \log n$ .

The normalised entropy criterion (NEC) was introduced by Celeux and Soromenho [11]; an improvement is due to Biernacki, Celeux, and Govaert [6]. This improved NEC chooses a model with s segments if  $NEC(s) \leq 1$ , ( $2 \leq s \leq S$ ) and states that  $NEC(1) = 1$ ; otherwise NEC declares there is no clustering structure in the data.

An approximate Bayesian solution, which is a crude approximation to twice the log Bayes factor for S segments [30], the approximate weight of evidence (AWE)

proposed by Banfield and Raftery [3], uses the classification likelihood, and chooses a model with  $S$  segments that minimises (3) with  $C = 2EN(S) + 2n_{\underline{\Psi}}(3/2 + \log n)$ .

Finally, the  $\mathcal{L}$  criterion [22] depends on sample size,  $n$ , number of model parameters,  $n_{\underline{\Psi}}$ , and the mixing probabilities,  $\lambda_s$ , and chooses a model with  $S$  segments that minimises  $-\text{LL} + (n_{\underline{\Psi}}/2)\sum \log(n\lambda_s/12) + S/2\log(n/12) + S(n_{\underline{\Psi}} + 1)/2$ .

### 3.2 Information criteria selection

In order to select one particular criterion for determining the number of segments in a LSM based on mixed type variables, we run some auxiliary clustering analysis. We analyse ten real data sets (table 3) with mixed variables (continuous and categorical) and known structure (the clusters in the data set are previously known) and use all the criteria presented on table 2, for the LSM estimation.

**Table 3.** The analyzed data sets

Data sets	source	Segmentation variables	Sample size	Data sets	source	Segmentation variables	Sample size
Bird survival	[23]	3 cont. 1 categ.	50	Imports-85	[27]	15 cont. 11 categ.	160
Methadone treatment	[23]	2 cont. 2 categ.	238	Heart	[17]	6 cont. 7 categ.	270
North Central Wisconsin	[15]	4 cont. 1 categ.	34	Cancer	[25]	8 cont. 4 categ.	471
Hepatitis	[20]	5 cont. 5 categ.	80	AIDS	[37]	3 cont. 3 categ.	944
Neolithic Tools	[16]	2 cont. 3 categ.	103	Ethylene Glycol	[33]	2 cont. 1 categ.	1028

Table 4 presents the proportion of data sets in which theoretical information criteria were able to recover the original cluster structure (in particular the true number of segments), as well as the corresponding criteria ranking.

**Table 4.** Information criteria performance

% of data sets where the true number of segments is identified of segments	Criteria Ranking
80	ICL-BIC
70	BIC and CAIC
60	$\mathcal{L}$
40	AIC, AICu, NEC, and AWE
30	AIC3
20	AICc
10	CLC

According to the obtained results ICL-BIC is the best performing criterion. It is able to recover the original data sets structure (it is able to detect the underlying true number of clusters or segments in the data set) in 8 of the 10 data sets (regardless the number of variables and sample size).

It is followed by CAIC and BIC (ex-aequo) and  $\mathcal{L}$  (3<sup>rd</sup> place). As a consequence, we opt for the use of the ICL-BIC criterion on the retail segmentation application (we also present results for BIC, CAIC, and  $\mathcal{L}$  criteria).

## Segmentation of Retail Clients

### 4.1 Data set description

The retail data set includes attributes referring to 1504 supermarket clients. Data originates from a questionnaire responses and includes several characteristics ranging from attitudes to demographics.

**Table 5.** Segmentation base variables

Segmentation Base	Type	Variables' Categories
<i>Amount spent on retail store</i>	Continuous	-
<i>Proportion of expenditure in retail chain</i>	Continuous	-
<i>Transportation</i>	Categorical	Car, Walking, Public Transportation, Motor cycle
<i>Usage frequency</i>	Categorical	Every day, 2 or 3 times a week, Once a week, Once a twice a week, ..., Occasional
<i>Visit pattern</i>	Categorical	During the week, During the weekend, Both situation
<i>Coming from</i>	Categorical	Home, Job, Passing by, Other
<i>Reasons for purchasing</i>	Categorical	Home proximity, job proximity, Passing by, Low prices, Brand diversity, ..., Doesn't know
<i>Travel time</i>	Categorical	2 minutes walking (mw), 2 to 5 mw, 5 to 10 mw, more than 10 mw, ..., more than 15 m by car, 10 to 15 m by car

As already referred, product-specific base variables are preferable for segmentation purposes. In order to segment retail clients we thus select some attitudinal and behavioural variables, such as *reasons to do the purchase*, *purchasing habits*, *usage frequency*, *visit pattern*, *travel time*, *amount spent* and *proportion of expenditure in the retail chain* (the proportion of monthly expenditures which refers to the specific supermarket chain). These variables illustrate the relationship between consumers and retail stores. Demographics such as *gender*, *age*, *income*, *occupation*, and *education* are available for identifying the individuals in the segments, turning them more accessible. Table 5 presents the segmentation base variables: 2 continuous and 6 categorical.



## 4.2 Segment Structure

Results from the estimation of a LSM using the referred segmentation variables (see Table 2) yield a two-segments structure. The ICL-BIC values corresponding to this and alternative solutions are displayed on table 6.

**Table 6.** Model selection

Number of segments	Number of parameters	BIC	CAIC	$\mathcal{L}$	ICL-BIC
1	76	56227,590	56303,59	28060,284	56227,589
2	119	55358,367	55477,367	27858,108	55980,600
3	162	55224,071	54896,180	28133,315	56219,539

Number of observations: 1504

Since we found there was an interaction between some base variables we included the following in the adopted model: usage frequency and visit pattern; transportation and travel time; amount spent and proportion of expenditure in the retail chain.

As we can see the ICL-BIC criterion attains its minimum for S=2 and  $\mathcal{L}$  yields the same conclusion.

**Table 7.** Segments and their characteristics

	Seg.1	Seg.2		Seg.1	Seg.2
<i>Usage Frequency</i>			<i>Travel time</i>		
Every day	41%	13%	Missing	2%	4%
2 or 3 times a week	35%	31%	2 minutes walking (mw)	21%	7%
Once a week	15%	21%	2 to 5 mw	26%	16%
Once or twice a week	2%	6%	5 to 10 mw	17%	13%
Once a month	6%	8%	More than 10 mw	7%	6%
Occasionally	1%	21%	Less or equal 5 m by car	16%	13%
<i>Reasons for purchasing</i>			5 to 10 m by car	5%	13%
Home proximity	75%	44%	10 to 15 m by car	4%	13%
Job proximity	3%	14%	More than 15 m by car	3%	14%
Passing by	2%	23%	<i>Visit pattern</i>		
Low prices	5%	3%	During the week	22%	40%
Brand diversity	1%	1%	During the weekend	12%	21%
Products diversity	2%	3%	Both situations	66%	38%
Habit	4%	4%	<i>Coming from</i>		
Product quality	3%	3%	Home	83%	51%
Fresh products' quality	1%	1%	Job	12%	35%
Store cleanliness	1%	1%	Passing by	3%	10%
Service promptness	1%	1%	Other	2%	4%
Service friendliness	2%	0%	<i>Transportation</i>		
Promotions	0%	1%	Car	25%	45%
Open hours	0%	1%	Walking	71%	43%
Parking conditions	0%	0%	Public transportation	4%	13%
Other reasons	0%	0%	Motorcycle	0%	0%
Doesn't know	1%	1%	-	-	-
<i>Proportion of expenditure in retail chain</i>			<i>Amount spent on retail store</i>		
Mean (€)	60,66	22,03	Mean (€)	342,5	255,9

We thus select a LSM with two segments (of sizes 917 for segment 1, and 587 for segment 2, by “modal allocation”), which we characterize on table 7.

As a result from the segments’ profiling we name Seg.1 as *Preferential clients* and Seg.2 as *Occasional clients*.

*Preferential clients* go often to the retail supermarkets; they leave nearby and walk to the super. These clients allocate 60% of their home monthly expenditures to the retail chain.

*Occasional clients* also include some clients that go often to the supermarket but they clearly differ from Seg.1 concerning the inclusion of occasional purchasers.

Location (Home proximity) is an important reason for purchasing for both segments; however for *Occasional clients*’ job proximity is also relevant. This segment also includes more clients which go to the super by car.

These results agree, in general, with those obtained in previous segmentation based on a larger sample and on a similar inquiry conducted two years before [10]. It is thus possible to conclude that this segment structure is stable.

## Conclusion and future work

In this article we discuss the use of Latent Segment Models for market segmentation. We focus on the utilization of theoretical information criteria to recover clustering structures. In particular, we discuss the use of these criteria for mixed type variables based clustering, since segmentation is typically based on attributes with diverse measurement levels. The discussion is motivated by an application: the segmentation of clients of a retail chain.

We first present the analysis of ten data sets with known clustering structure and rank several criteria according to their ability to recover the original structure, indicating the correct number of clusters. According to the obtained results we rank the best criteria as follows: ICL-BIC (1st place), BIC and CAIC (2nd place, ex-aequo) and  $\mathcal{L}$  (3rd place). Using this empirical test’s results we select the ICL-BIC (a criterion which was specifically designed for clustering applications) criterion as an indicator of the correct number of retail clients’ latent segments.

We finally estimate a Latent Segments Model to obtain a segment structure which refers to the clients of a supermarkets retail chain. We use product specific variables as a base for segmentation (e.g. reasons for purchase). As a result (and using the ICL-BIC criterion) two segments are constituted. The Preferential clients segment and the Occasional Clients segment.

In addition to these substantive conclusions we consider that the issue concerning the selection of specific information criteria to estimate Latent Segments Models, based on mixed type data, should be further discussed. In fact, mixed type variables are commonly considered in segmentation studies and thus, the information criteria performance which is empirically observed in the present work deserves future research. In the present work empirical results provide a criteria ranking. Naturally, a larger amount of data sets with diverse characteristics (which may be obtained via simulation procedures) should be considered in order to further prove the consistency of the present conclusions.

## References

1. H. Akaike, Information Theory and an Extension of Maximum Likelihood Principle, in K. T. Emanuel Parzen, Genshiro Kitagawa, ed., Selected Papers of Hirotugu Akaike, in Proceedings of the Second International Symposium on Information Theory, B.N. Petrov and F. Casati, eds., Akademiai Kiado, Budapest, 1973, 267-281, Springer-Verlag New York, Inc, Texas, 1973, pp. 434.
2. H. Akaike, Maximum likelihood identification of Gaussian autoregressive moving average models, *Biometrika*, 60 (1973), pp. 255-265.
3. J. D. Banfield and A. E. Raftery, Model-Based Gaussian and Non-Gaussian Clustering, *Biometrics*, 49 (1993), pp. 803-821.
4. C. Biernacki, Choix de modèles en Classification-Ph.D. Tesis., 1997.
5. C. Biernacki, G. Celeux and G. Govaert, Assessing a Mixture model for Clustering with the integrated Completed Likelihood, *IEEE Transactions on Pattern analysis and Machine Intelligence*, 22 (2000), pp. 719-725.
6. C. Biernacki, G. Celeux and G. Govaert, An improvement of the NEC criterion for assessing the number of clusters in mixture model, *Pattern Recognition Letters*, 20 (1999), pp. 267-272.
7. C. Biernacki and G. Govaert, Using the classification likelihood to choose the number of clusters, *Computing Science and Statistics*, 29 (1997), pp. 451-457.
8. H. Bozdogan, Mixture-Model Cluster Analysis using Model Selection criteria and a new Informational Measure of Complexity, in H. Bozdogan, ed., Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Approach, 69-113, Kluwer Academic Publishers, 1994, pp. 69-113.
9. H. Bozdogan, Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions, *Psychometrika*, 52 (1987), pp. 345-370.
10. M. G. M. S. Cardoso and A. B. Mendes, Segmentação de clientes de lojas de Pequena Dimensão, IX Congresso da Sociedade Portuguesa de Estatística, Ponta Delgada, 2002, pp. 157-170.
11. G. Celeux and G. Soromenho, An entropy criterion for accessing the number of clusters in a mixture model, *Journal of Classification*, 13 (1996), pp. 195-212.
12. J. Y. Ching, A. K. C. Wong and K. C. C. Chan, Class-Dependent Discretization for Inductive Learning from Continuous and Mixed-Mode Data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17 (1995), pp. 641-651.
13. Y.-S. Choi, B.-R. Moon and S. Y. Seo, Proceedings of the 2005 Conference on Genetic Fuzzy Discretization with Adaptive Intervals for Classification Problems, Conference on Genetic and Evolutionary Computation, ACM Press, New York, NT, USA, Washington DC, USA, 2005, pp. 2037-2043.
14. S. H. Cohen and V. Ramaswamy, Latent segmentation models, *Marketing Research Magazine*, Summer (1998), pp. 15-22.
15. S. L. Crawford, M. H. Degroot and J. B. Kadane, Modeling Lake-Chemistry Distributions: Approximate Bayesian Methods for Estimating a Finite-Mixture Model, *Technometrics*, 34 (1992), pp. 441-453.
16. P. Dellaportas, Bayesian Classification of Neolithic Tools, *Applied Statistics*, 47 (1998), pp. 279-297.
17. R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, K. Schmid, K. Sandhu, K. Guppy, S. Lee and V. Froelicher, Rapid searches for complex patterns in biological molecules, *American Journal of Cardiology*, 64 (1989), pp. 304-310.
18. J. G. Dias, Finite Mixture Models; Review, Applications, and Computer-intensive Methods, Economics, Groningen University, PhD Thesis, Groningen, 2004, pp. 199.

19. W. R. Dillon and A. Kumar, Latent structure and other mixture models in marketing: An integrative survey and overview, chapter 9 in R.P. Bagozi (ed.), *Advanced methods of Marketing Research*, 352-388, Cambridge: blackwell Publishers, 1994.
20. B. Efron, and Gong, G., A Leisurely at the Bootstrap, the Jackknife, and cross-Validation, *The American Statistician*, 37 (1983), pp. 36-48.
21. G. Fennell, G. M. Allenby, S. Yang and Y. Edwards, The effectiveness of Demographic and Psychographic Variables for Explaining Brand and Product Category Use, *Quantitative Marketing and Economics*, 1 (2003), pp. 233-244.
22. M. A. T. Figueiredo and A. K. Jain, Unsupervised Learning of Finite Mixture Models, *IEEE Transactions on pattern analysis and Machine Intelligence*, 24 (2002), pp. 1-16.
23. D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway and E. Ostrowski, *Small Data Sets*, Chapman & Hall, London, 1996.
24. L. Hunt and M. Jorgensen, Mixture model clustering for mixed data with missing information, *Computational Statistics & Data Analysis*, 41 (2003), pp. 429-440.
25. L. A. Hunt and K. E. Basford, Fitting a Mixture Model to Three-Mode Three-Way Data with Categorical and Continuous Variables, *Journal of Classification*, 16 (1999), pp. 283-296.
26. C. M. Hurvich and C.-L. Tsai, regression and Time Series Model Selection in Small Samples, *Biometrika*, 76 (1989), pp. 297-307.
27. D. Kibler, D. W. Aha and M. Albert, Instance-based prediction of real-valued attributes, *Computational Intelligence*, 5 (1989).
28. B.-D. Kim, K. Srinivasan and R. T. Wilcox, Identifying Price Sensitive Consumers: The Relative Merits of Demographic vs. Purchase Pattern Information, *Journal of Retailing*, 75 (1999), pp. 173-193.
29. B. G. Leroux and M. L. Puterman, Maximum-Penalized-Likelihood Estimation for Independent and Markov-Dependent Mixture Models, *Biometrics*, 48 (1992), pp. 545-558.
30. G. F. McLachlan and D. Peel, *Finite Mixture Models*, John Wiley & Sons, Inc., 2000.
31. A. McQuarrie, R. Shumway and C.-L. Tsai, The model selection criterion AIC<sub>u</sub>, *Statistics & Probability Letters*, 34 (1997), pp. 285-292.
32. I. Moustaki and I. Papageorgiou, Latent class models for mixed variables with applications in Archaeometry, *Computational Statistics & Data Analysis*, In Press (2004).
33. C. J. Price, C. A. Kimmel, R. W. Tyl and M. C. Marr, The development toxicity of ethylene glycol in rats and mice, *Toxicological Applications in Pharmacology*, 81 (1985), pp. 113-127.
34. R. A. Redner and H. F. Walker, Mixture Densities, Maximum Likelihood and the EM Algorithm, *SIAM review*, 26 (1984), pp. 195-239.
35. J. Rissanen, Modeling by shortest data description, *Automatica*, 14 (1978), pp. 465-471.
36. G. Schwarz, Estimating the Dimension of a Model, *The Annals of Statistics*, 6 (1978), pp. 461-464.
37. J.-Q. Shi and S.-Y. Lee, Latent variable models with mixed continuous and polytomous data, *Journal of the Royal Statistical Society, Series A*, 62 (2000).
38. W. R. Smith, Product differentiation and market segmentation as alternative marketing strategies, *Journal of Marketing*, 21 (1956), pp. 3-8.
39. C.-Y. Tsai and C.-C. Chiu, A purchase-based market segmentation methodology, *Expert Systems with Applications*, 27 (2004), pp. 265-276.
40. J. K. Vermunt and J. Magidson, Latent class cluster analysis., J.A. Hagenaars and A.L. McCutcheon (eds.), *Applied Latent Class Analysis*, 89-106., Cambridge University Press, 2002.
41. M. Vriens, Market Segmentation. Analytical Developments and Applications Guidelines, Technical Overview Series, Millward Brown IntelliQuest, 2001, pp. 1-42.
42. M. Wedel and W. A. Kamakura, *Market Segmentation: Concepts and methodological foundations*, Kluwer Academic Publishers, Boston, 1998.