13. Mixture models

Jeroen K. Vermunt and Leo J. Paas

13.1. INTRODUCTION

Marketing research literature, and most quantitative literature in other fields, addresses two main topics: (1) Categorization and (2) Prediction. Categorization can involve the following types of units: variables, companies, countries, consumers, customers, brands, websites, etc. For example, variables can be categorized using factor analysis or other techniques to assess the dimensionality underlying response patterns to the observed items (Paas and Sijtsma, 2008). Categorization can also assess which brands may be competing for the same market of consumers and which brands can be considered as non-competitive, by allocating brands in groups (Rossiter and Bellman, 2005). A very common categorization in the marketing research literature concerns the allocation of consumers or clients to different segments, where needs and wants differ across segments (Wedel and Kamakura, 2000) and result in segment-specific strategies. This will be the emphasis of the current contribution.

Consumer and client-base segmentation have received much attention (Wedel and Kamakura, 2000) since the basic concepts were introduced by Smith (1956). Traditional heuristic-based clustering procedures have been used extensively for deriving segments from databases on clients or consumers. However, since the first application of latent class analysis, a type of mixture model, was reported in a marketing journal (Green et al., 1976), the model-based mixture modeling approach has been the preferred method for market segmentation (Leeflang et al., 2000). Interestingly, the marketing literature was early in adopting mixture modeling, two years after the seminal publication of the Goodman (1974) and even one year before the publication of the article on the EM-algorithm (Dempster et al, 1977), which made estimation of mixture models more feasible.

Mixture models are highly flexible, enabling the modeling of data with various types of structures (Leeflang et al., 2000). Table 13.1 summarizes the main types of data structures with the corresponding type of mixture model. Basically, we distinguish between data sets with either a single or multiple columns for the response variable(s), and with a single or multiple rows per observational unit, where for the latter we make a further distinction between an arbitrary and a meaningful (longitudinal) ordering of the rows. Simple mixture models are typically used only for density estimation and are of less interest for marketing applications. Mixture regression and mixture growth models are random-effects like models: the aim is to describe the heterogeneity in the regression or growth parameters by assuming that individuals belong to finite number of latent classes. These models will always contain predictors affecting the responses (in mixture growth models these are time-variables). In the second row, we see the more typical cluster analysis like applications of LC analysis. Here one may also include predictors affecting the responses, yielding multivariate extensions of mixture regression and growth models.

Table 13.1: Typology of mixture models based other type of data structure

	Independent observations	Dependent observations or multilevel data	Longitudinal or panel data
Univariate response	(1) simple mixture model	(2) mixture regression model	(3) mixture growth model
Multivariate responses	(4) LC model	(5) multilevel LC model	(6) (mixture) latent Markov model

In this chapter we further discuss the model types 4 and 5 in Table 13.1, starting with the basic idea of a simple mixture model (category 1 in Table 13.1). After that we focus on models for multivariate responses, which correspond to cluster-analysis like applications of mixture models, i.e., the standard latent class model (category 4), with covariates, as well as its multilevel extension for situations in which lower-level observations (e.g. individuals) are nested within higher-level units (e.g. regions), and in which we wish to cluster/segment both lower- and higher-level units, i.e., the multilevel LC model (category 5). Note that, (mixture) latent Markov models (category 6), which can be used for studying how individuals move between latent states, clusters, or segments over time, have been discussed previously in this series by Leeflang et al. (2000), and will also be addressed in detail in the next chapter of the current book. We illustrate applications of the standard and multilevel LC models based on applications in international marketing. We end with a discussion of statistical software implementing these models and a brief summary, which also introduces other marketing applications of the LC model and the multilevel LC model.

13.2. MIXTURE MODELS

13.2.1. The simple mixture model

The basic assumption of any type of mixture model is that the population consists of a finite number of unobserved groups which differ with respect to the parameters of a statistical model. These unobserved groups are referred to as mixture components or latent classes. The assumed statistical model within latent classes is typically rather simple; for example, a Poisson distribution for a count variable, a linear regression model with normal errors for a continuous outcome variables, or an independence model for a set of categorical variables.

Let us first look at a simple mixture model (category 1 in Table 13.1) , say a count variable denoted by y_i, where i refers to one of the individuals in the sample and $1 \leq i \leq N$. The aim of the analysis is to describe the sample distribution of this variable assuming 1) there are *C* latent classes and 2) y_i follows a Poisson distribution with mean μ_c within the classes, where *c* refers to a particular latent class. The corresponding mixture model can be expressed as follows:

$$P(y_i) = \sum_{c=1}^{C} \pi_c \, P(y_i|c) = \sum_{c=1}^{C} \pi_c \frac{e^{-\mu_c} {\mu_c}^{y_i}}{y_i!}. \quad (13.1)$$

Here, π_c is the class proportion or the probability of belonging to class *c*, where $\pi_c > 0$ and $\sum_{c=1}^{C} \pi_c = 1$. As can be seen, the probability of observing a count value of y_i, $P(y_i)$, is assumed to be a weighted average of the class-specific probabilities $P(y_i|c)$, where class proportions are weights (McLachlan and Peel, 2000). The $P(y_i|c)$ are Poisson distributions with means μ_c.

Table 13.2 illustrates application of this model. It presents the observed frequency distribution of a count variable as well as its estimated frequency distributions under a standard and three-class Poisson model. As can be seen, while the standard Poisson model does not fit the data, the observed distribution is approximated very well with three latent classes. This shows a general result: complex observed distributions can typically be approximated well using a mixture of simple distributions, which is why mixture models are well suited for density estimation.

Table 13.2: Observed and estimated frequency distributions under a standard and 3-class mixture Poisson model

Count	Observed	Standard Poisson	3-Class Poisson
0	102	8.43	101.99
1	54	33.63	53.93
2	49	67.11	50.20
3	62	89.28	54.23
4	44	89.09	47.42
5	25	71.11	34.14
6	26	47.30	21.93
7	15	26.97	14.28
8	15	13.46	10.95
9	10	5.97	10.13
10	10	2.38	10.14
11	10	0.86	9.95
12	10	0.29	9.19
13	3	0.09	7.90
14	3	0.03	6.32
15	5	0.01	4.72
16	5	0.00	3.31
17	4	0.00	2.18
18	1	0.00	1.36
19	2	0.00	0.80
20	1	0.00	0.45

Note: In the standard model, the Poisson rate equals 3.99. In the 3-class model, the class-specific rates are .29, 3.48, and 11.22, and the class proportions are .28, .54, and .18, respectively.

13.2.2. The unrestricted LC model for multivariate responses

In the LC model (category 4), which can be considered as cluster- or segmentation-like applications of mixture models, we will typically have multiple responses per individual. We denote these responses by y_{ik}, where k denotes a particular response variable and K the number of response variables ($1 \leq k \leq K$). Denoting the full response vector of person i as $\mathbf{y}_i$, a LC model can be defined as follows (Lazarsfeld, 1950; Goodman, 1974):

$$P(\mathbf{y}_i) = \sum_{c=1}^{C} \pi_c \, P(\mathbf{y}_i|c) = \sum_{c=1}^{C} \pi_c \prod_{k=1}^{K} P(y_{ik}|c). \quad (13.2)$$

As can be seen, similar to the simple mixture model presented above, we assume $P(\mathbf{y}_i)$ to be a mixture of class-specific distributions $P(\mathbf{y}_i|c)$. However, here we make a second important assumption, namely that the *K* responses are independent of one another conditional on a person's class membership. That is, we assume $P(\mathbf{y}_i|c) = \prod_{k=1}^{K} P(y_{ik}|c)$. This is usually referred to as the local independence assumption. This assumption is also used in other types of latent variable models, such as in factor analysis and Item Response Theory (IRT) modeling, as well as in random-effects models, meaning that it is not specific for LC analysis. Note that IRT models will be discussed further in section 13.2.3. As shown by Hagenaars (1988) and Oberski, van Kollenburg, and Vermunt (2013), this assumption can be tested and also be relaxed for some item pairs.

The specific form used for the class-specific probability density $P(y_{ik}|c)$ depends on the scale type of y_{ik}: Examples include a normal distribution for continuous y_{ik}, a Poisson distribution for count y_{ik}, a Bernoulli distribution for dichotomous y_{ik}, and a multinomial distribution for polytomous y_{ik}. Moreover, different forms can be combined within the same model (Vermunt and Magidson, 2002). Let us look in more detail at the Bernoulli case, in which $P(y_{ik}|c) = \pi_{kc}{}^{y_{ik}}(1 - \pi_{kc})^{(1-y_{ik})}$, meaning that the class-specific parameters are the success probabilities π_{kc}. Table 13.3 presents a small illustrative data set consisting of three dichotomous responses from a hypothetical customer survey asking whether respondents own three types of electronic devices. Table 13.4 reports the parameter estimates (class proportions π_c and class-specific success probabilitiesπ_{kc}) obtained with a 2-class model. For a subject belonging to the first latent class, the ownership probabilities equal .844, .912, and .730 for products 1, 2, and 3, respectively. The local independence assumption implies, for example, that the probability of owning only the first two products equals .844×.912× (1-.730) = 0.208 in Latent Class one, and 0.091 in Latent Class two. Furthermore, a LC model defines the *overall* probability of having a particular response pattern, which turns out to be 0.161 for the owning of products 1 and 2. The latter number is obtained as a weighted average of the class-specific joint response probabilities taking into account the class proportions (.601 and .399); that is, .601×0.208+.399×0.091=0.161. The LC model was used in the first marketing application of mixture modeling reported by Green et al. (1976).

Table 13.3: Small data set with three dichotomous responses

y_{i1}	y_{i2}	y_{i3}	Frequency	$P(\mathbf{y}_i \mid c=1)$	$P(\mathbf{y}_i \mid c=2)$	$P(\mathbf{y}_i)$	$P(c=1 \mid \mathbf{y}_i)$	$P(c=1 \mid \mathbf{y}_i)$	Modal
0	0	0	239	0.004	0.272	0.111	0.020	0.980	2
0	0	1	101	0.010	0.102	0.047	0.128	0.872	2
0	1	0	283	0.038	0.271	0.131	0.175	0.825	2
0	1	1	222	0.104	0.102	0.103	0.605	0.395	1
1	0	0	105	0.020	0.092	0.049	0.248	0.753	2
1	0	1	100	0.054	0.035	0.046	0.703	0.297	1
1	1	0	348	0.208	0.091	0.161	0.774	0.226	1
1	1	1	758	0.562	0.034	0.352	0.961	0.039	1

Table 13.4: Parameters (class proportions and ownership probabilities) obtained with a 2-class model for the data in Table 13.3

Parameter	Class 1	Class 2
π_c	0.601	0.399
π_{1c}	0.844	0.252
π_{2c}	0.912	0.499
π_{3c}	0.730	0.273

Similar to cluster analysis, one of the purposes of the LC model might be to assign individuals to latent classes (LCs). The probability of belonging to LC *c* given responses $\mathbf{y}_i$ – often referred to as the posterior class membership probability – can be obtained by the Bayes rule:

$$P(c \mid \mathbf{y}_i) = \frac{\pi_c P(\mathbf{y}_i \mid c)}{P(\mathbf{y}_i)}. \quad (13.3)$$

Table 13.2 reports $P(c \mid \mathbf{y}_i)$ for each answer pattern. For example, $P(c=1 \mid \mathbf{y}_i)$ equals 0.774 for the (1,1,0) pattern, which is obtained as .601×.208/.161. The most common classification rule is modal assignment, which amounts to assigning each individual to the LC for which $P(c \mid \mathbf{y}_i)$ is largest. The last column of Table 3 reports the modal assignments and shows that consumers that own at least 2 products are assigned to class 1 and the others to class 2.

The assigned class membership can among others be used to investigate the relationship between class-membership and external variables (for example, concomitant variables). Such variables can however also be included in the LC model itself. We discuss this in more detail in section 13.2.6.

Table 13.5: Parameters (class proportions and ownership probabilities) obtained with a Proctor model for the data in Table 13.3

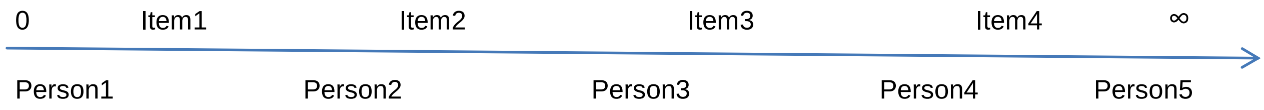
Parameter	Class 1	Class 2	Class 3	Class 4
π_c	0.160	0.155	0.126	0.559
π_{1c}	0.167	0.833	0.833	0.833
π_{2c}	0.167	0.167	0.833	0.833
π_{3c}	0.167	0.167	0.167	0.833

13.2.3. Some restricted LC models for categorical responses

Interesting types of restricted LC models for categorical items have been proposed, which involve imposing (linear) constraints on either the conditional probabilities or the corresponding logit coefficients. Of particular interest are probabilistic Guttman scaling models for dichotomous responses, which are LC models with $C=K+1$ classes, in datasets with K items. Various marketing papers presented applications of such restricted LC models, e.g., Bijmolt el. (2004), Feick (1987) and Paas and Molenaar (2005).

The basic rationale on which Guttman scaling is based can be explained using Figure 1, which is based on a hypothetical dataset in which five persons reacted to four items, e.g., survey questions. The line represents the underlying latent trait on which both persons and items are ordered. A latent trait could be brand knowledge, with low values representing a low level of knowledge and higher positions on the latent trait representing much knowledge on the focal brand. Person1 has the lowest position on the latent trait, a position below all items, implying this person possesses too little knowledge to answer any of the items correctly. Person2 is found between item1 and item2, implying Person2 will probably answer item1 correctly, but not the four other items. Person3's knowledge is likely to lead to correct answers on items 1 and 2, but not on the other items. Moving all the way up the latent trait will lead to person5, who possesses sufficient brand knowledge for answering all items correctly.

Figure 1: Hypothetical Guttman scale



The hierarchical ordering of items and persons implies that if a person answers a difficult brand knowledge item correctly, such as item 4, this person should also answer the easier items correctly, items 1 to 3 in this example. Stated more formally and in terms of LC models, apart from measurement error, class *c* should provide a positive answer to the *c*-1 easiest items and a negative answer to the remaining *K*-(*c*-1) items. This rather basic Guttman scale model forms the fundamentals for various types of IRT models. Most relevant herein is that the answer patterns of respondents will not be deterministic, some answers maybe be inconsistent with the hierarchical ordering of items. In our hypothetical example a Guttman error occurs if a respondent answers the more difficult item4 correctly but not the easier item2. Probabilistic IRT models accommodate for the occurrence of such Guttmann errors in various ways. Refer to Hambleton and Swaminathan (1985) or Sijtsma and Molenaar (2002) for a more formal and elaborate description of item response theory. These different models can be represented in a LC model framework, as we will discuss briefly in this section.

The various types of probabilistic Guttman models differ in the constraints they impose on the measurement error. The simplest and most restricted model is the Proctor (1970) model. Table 13.5 presents the parameter estimates obtained when fitting the Proctor model to the data set in Table 13.3. As can be seen, the ownership probability is either .833 or .167=1-.833. The measurement error – or the probability of owning a product which is not in agreement with the class – is estimated to be equal to .167. Whereas the Proctor model assumes that the measurement error is constant across items and classes, less restricted models can be defined, with different error probabilities across items, classes, or both (Dayton, 1999).

Croon (1990) proposed a restricted LC model that similarly to non-parametric IRT (Sijtsma and Molenaar, 2000) assumes monotonic item response functions; that is, $\pi_{kc} \leq \pi_{k,c+1}$. A restricted version, in which not only classes but also items are ordered, is obtained by imposing the additional set of restrictions, $\pi_{kc} \leq \pi_{k+1.c}$; that is, by assuming double monotony. Vermunt (2001) discussed various generalizations of these models. In the marketing literature such model restrictions were applied for predicting consumer's future product acquisition using cross sectional data on ownership of financial products (Paas and Molenaar, 2005).

Various authors described the connection between restricted LC models and parametric IRT modeling (see, for example, Heinen, 1996; Lindsay, Clogg, and Grego, 1991); that is, IRT models with a discrete specification of the distribution of the underlying trait or ability can be defined as LC models in which the response probabilities are parametrized using logistic equations with constraints on the logit parameters, where LC locations represent the *C* possible values of the discretized latent trait. These locations may be fixed a priori, for example, at -2, -1, 0, 1, and 2 in the case of *C*=5, but may also be treated as free parameters to be estimated. Depending on whether the items are dichotomous, ordinal, or nominal, this yields a 2-parameter logistic, generalized partial credit, or nominal response model. Further restrictions involve equating slope parameters across items, yielding Rasch and partial credit models, and imposing across category and across item restrictions on intercept parameters as in rating scale models for ordinal items. Multidimensional variants can be obtained using what Magidson and Vermunt (2001) refer to as discrete factor models.

13.2.4. Parameter estimation by maximum likelihood

The parameters of LC models are typically estimated by means of maximum likelihood (ML). The log-likelihood function that is maximized is based on the probability density $P(\mathbf{y}_i)$, that is,

$$\ln L = \sum_{i=1}^{N} \ln P(\mathbf{y}_i). \tag{13.4}$$

With categorical responses one will typically group the data and construct a frequency table as we did in Table 3. The log-likelihood function for grouped data equals

$$\ln L = \sum_{m=1}^{M} n_m \ln P(\mathbf{y}_m). \tag{13.5}$$

where m is a data pattern, M the number of different data patterns, and n_m the cell count corresponding to data pattern m. Notice that only nonzero observed cell entries contribute to the log-likelihood function, a feature that is exploited by the more efficient LC software packages that have been developed over the last decades. These packages will be discussed in section 13.4.

One of the problems in the estimation of LC models for discrete y_{ik} is that model parameters may be nonidentified, even if the number of degrees of freedom – the number of independent cells in the K-way cross-tabulation minus the number of free parameters – is larger than or equal to zero. Non-identification means that different sets of parameter values yield the same maximum of the log-likelihood function or, worded differently, that there is no unique set of parameter estimates. The formal identification check is via the Jacobian matrix (matrix of first derivatives of $P(\mathbf{y}_i)$), which should be column full rank. Another option is to estimate the model of interest with different sets of starting values. Except for local solutions (see below), an identified model gives the same final estimates for each set of the starting values.

Although there are no general rules with respect to the identification of LC models, it is possible to provide certain minimal requirements and point to possible pitfalls. For an unrestricted LC model, one needs at least three responses (y_{ik}'s) per individual, but if these are dichotomous, no more than two latent classes can be identified. Consideration is required when analyzing four dichotomous response variables, in which case the unrestricted three-class model is not identified, even though it has a positive number of degrees of freedom. With five dichotomous items, however, even a five-class model is identified. Usually, it is possible to achieve identification by constraining model parameters.

A second problem associated with the estimation of LC models is the presence of local maxima. The log-likelihood function of a LC model is not always concave, which means that hill-climbing algorithms may converge to a different maximum depending on the starting values. Usually, we are looking for the global maximum. The best way to proceed is, therefore, to estimate the model with different sets of random starting values. Typically, several sets converge to the same highest log-likelihood value, which can then be assumed to be the ML solution. Some software packages have automated the use of multiple sets of random starting values to reduce the probability of getting a local solution.

Another problem in LC models is the occurrence of boundary solutions, which are probabilities equal to 0 (or 1) or logit parameters equal to minus (or plus) infinity. These may cause numerical problems in the estimation algorithms, occurrence of local solutions, and complications in the computation of standard errors and number of degrees of freedom of the goodness-of-fit tests. Boundary solutions can be prevented by imposing constraints or by taking into account other kinds of prior information on the model parameters.

The most popular methods for solving the ML estimation problem are the expectation-maximization (EM) and Newton-Raphson (NR) algorithms. EM is a very stable iterative method for ML estimation with incomplete data (Dempster et al, 1977). NR is a faster procedure that, however, needs good starting values to converge. The latter method makes use of the matrix of second-order derivatives of the log-likelihood function, which is also needed for obtaining standard errors of the model parameters.

13.2.5. Model selection issues

A challenging and fundamental decision for LC models and other categories of models, in Table 13.1, concerns model specification. With most available data sets model specifications can differ in various ways. Models may differ in the number of latent classes, the specification of the measurement model, freely estimated or restricted as discussed in section 13.2.4, the relationship between the measured indicators and the latent classes, nominal, ordinal or metric, specification of the form of the covariate effects, nominal, ordinal or metric, etc.

The goodness-of-fit of such alternative formulation of LC models for categorical responses can be tested using Pearson and likelihood-ratio chi-squared tests. The latter is defined as

$$G^2 = 2\sum_{m=1}^{M} n_m \ln \frac{n_m}{N\cdot P(\mathbf{y}_m)}. \qquad (13.6)$$

As in log-linear analysis, the number of degrees of freedom (*df*) equals the number of cells in the frequency table minus 1, minus the number of independent parameters ($npar$). In an unrestricted LC model,

$$df = \prod_{k=1}^{K} R_k - 1 - npar = \prod_{k=1}^{K} R_k - C\cdot[1 + \sum_{k=1}^{K}(R_k - 1)], \qquad (13.7)$$

where R_k is the number of categories of the k^{th} response variable. Although it is no problem to estimate LC models with 10, 20, or 50 indicators, in such cases, the frequency table may become very sparse and, as a result, asymptotic *p*-values can no longer be trusted. An elegant but time-consuming solution to this problem is to estimate the *p*-values by parametric bootstrapping. This procedure constructs the sampling distribution of the statistic of interested using Monte-Carlo simulation. More specifically, one generates M samples from the population defined by the estimated parameters and estimates the model with each of these M samples. The p-value is the proportion of samples in which the statistic is larger than in the original sample.

Another alternative for assessing model fit in sparse tables is to look at the fit in lower-order marginal tables (e.g., in the two-way marginal tables). An example is the bivariate residual, which is a Pearson chi-squared statistic for a two-way table divided by the number of degrees of freedom (Oberski, van Kollenburg, and Vermunt, 2013); that is,

$$BVR_{kk'} = \sum_{r=1}^{R_k} \sum_{r'=1}^{R_{k'}} \frac{[n_{kk'(rr')} - N \cdot P(y_k = r, y_{k'} = r')]^2}{N \cdot P(y_k = r, y_{k'} = r')} / [(R_k - 1)(R_{k'} - 1)]. \quad (13.8)$$

Here $n_{kk'(rr')}$ is the observed number of persons with responses r and r' on item pair k and k', and $P(y_k = r, y_{k'} = r')$ is the probability of this response pattern according to the estimated LC model. It can be computed by

$$P(y_k = r, y_{k'} = r') = \sum_{c=1}^{C} \pi_c P(y_k = r|c) P(y_{k'} = r'|c), \quad (13.9)$$

thus basically by applying the LC formula to an item pair. Larger BVR values point at possible violations of the local independence assumption for the item pairs concerned.

Even though LC models with *C* and *C* + 1 classes are nested, one cannot test them against each other using a standard likelihood-ratio test because it does not have an asymptotic chi-squared distribution. A way out to this problem is to approximate its sampling distribution using bootstrapping. But since this bootstrap likelihood-ratio method is computationally demanding, alternative methods are often used for comparing models with different numbers of classes. Most popular are information criteria, e.g., the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC), and the AIC3 measure, defined as: $BIC = -2\ln L + \ln(N)\,npar$, $AIC = -2\ln L + 2npar$, and $AIC3 = -2\ln L + 3\,npar$. Lower values imply better fit and parsimony.

Usually, we are not only interested in (relative) goodness-of-fit but also in how well class memberships can be predicted from the observed responses or, worded differently, how well classes are separated. This is can among other be quantified based on the estimated proportion of classification errors under modal classification, which equals

$$CE = \sum_{i=1}^{N} \frac{1}{N}[1 - \max P(c|\mathbf{y}_i)] \tag{13.10}$$

This number can be compared to the proportion of classification errors based on the unconditional probabilities π_c, yielding a reduction of errors measure called Lambda:

$$\text{Lambda} = \sum_{i=1}^{n} 1 - \frac{CE}{1-\max(\pi_c)} \tag{13.11}$$

The closer this nominal pseudo R^2 measure is to 1, the better the model performs in terms of classification accuracy. Other types of R^2 measures have been proposed based on entropy and qualitative variance; that is, using $\sum_{i=1}^{n}\sum_{c=1}^{C} - P(c|\mathbf{y}_i)\ln P(c|\mathbf{y}_i)\,/N$ and $\sum_{i=1}^{n}[1 - \sum_{c=1}^{C} P(c|\mathbf{y}_i)^2]/N$ as measures for class separation.

Finally, there is a class of measures which are similar to information criteria, but which also take into account classification performance. In other words, these measures try to balance fit, parsimony, and classification performance. The best-known of these measures are AWE and ICL-BIC.

13.2.6. LC analysis with concomitant variables

An important extension of the LC model involves inclusion of concomitant variables or covariates predicting class membership (Dayton and Macready, 1988; Kamakura, Wedel and Agrawal, 1994). Denoting a person's covariate vector by $\mathbf{x}_i$, this extended LC model is defined as:

$$P(\mathbf{y}_i|\mathbf{x}_i) = \sum_{c=1}^{C} \pi_{c|\mathbf{x}_i} P(\mathbf{y}_i|c) = \sum_{c=1}^{C} \pi_{c|\mathbf{x}_i} \prod_{k=1}^{K} P(y_{ik}|c). \tag{13.12}$$

The main change compared to the LC model is that the class membership probabilities may now be dependent on $\mathbf{x}_i$, whereas the definition of $P(\mathbf{y}_i|c)$ remains unchanged. Note that an additional assumption is made, namely that the effect of the $\mathbf{x}_i$ on the $\mathbf{y}_i$ is fully mediated by the latent classes. Section 13.3.1 provides an example of such a LC model, in which country of residence is the covariate of interest. It is possible to test this assumption using local fit measures similar to those discussed earlier, as well as to relax it by allowing for direct effects, which implies replacing $P(y_{ik}|c)$ by $P(y_{ik}|c, \mathbf{x}_i)$ for one or more of the y_{ik}.

Typically, $\pi_{c|\mathbf{x}_i}$ is modeled using a multinomial logistic specification; that is,

$$\pi_{c|\mathbf{x}_i} = \frac{\exp(+\sum_{p=1}^{P} \gamma_{pc} x_{ip})}{\sum_{c'=1}^{C} \exp(\gamma_{0c'} + \sum_{p=1}^{P} \gamma_{pc'} x_{ip})}, \quad (13.13)$$

Where γ_{0c} represents the intercept and γ_{pc} the slope for predictor x_{ip} for latent class c. For identification, we will typically assume parameters to sum to 0 across classes (effect coding) or to be equal to 0 for one class (dummy coding).

The simultaneous modeling of responses $\mathbf{y}_i$ and concomitant variables $\mathbf{x}_i$ may sometimes be impractical, especially when the number of possibly relevant concomitant variables is large. Therefore, researchers often prefer using a three-step analysis approach. This involves:

1) estimating a LC model without covariates,
2) obtaining the individuals' class assignment using the posterior membership probabilities, and
3) investigating how the class assignments are related to covariates.

The advantage of this approach is that it does not lead to a segmentation structure that is determined by the chosen covariates, only indicator variables influence this. This is particularly relevant when there is little theory on the relationship between covariates and latent class membership and possible direct effects of covariates on responses, which is often the case in the more or less exploratory clustering analyses conducted in marketing research.

However, as shown by Bolck, Croon, and Hagenaars (2004), this yields downward biased estimates of the covariate effects. Based on the work of these authors, Vermunt (2010) proposed a simple method to adjust for this bias (see also Bakk, Tekle, and Vermunt, 2013). The adjustment is based on the following relationship between the class assignments w_i and the true class membership *c:*

$$P(w_i|\mathbf{x}_i) = \sum_{c=1}^{C} \pi_{c|\mathbf{x}_i} P(w_i|c). \quad (13.14)$$

Note that this is again a LC model, but with w_i as a single "response" variable. The adjustment proposed by Vermunt (2010) therefore involves estimating a LC model with $\mathbf{x}_i$ as concomitant variables and w_i as the single response variable, while fixing the $P(w_i|c)$ at the values computed using the parameter estimates from the first step.

13.2.7. Multilevel LC analysis

Another important extension of the LC model concerns its adaptation for the analysis of multilevel data sets (Vermunt, 2003). Description of this multilevel extension of the LC model requires expansion of our notation. We refer to a particular higher-level unit as j and to the response vector of a group and an individual within a group as $\mathbf{y}_j$ and $\mathbf{y}_{ji}$, respectively. The number of individuals within a group is denoted by n_j, a group-level latent class or cluster by d, and the number of group-level clusters by D. The lower-level part of the multi-level LC model (category 5), a two-level LC model in this case, has the following form:

$$P(\mathbf{y}_{ji}|d) = \sum_{c=1}^{C} \pi_{c|d} P(\mathbf{y}_{ji}|c) = \sum_{c=1}^{C} \pi_{c|d} \prod_{k=1}^{K} P(y_{jik}|c), \quad (13.15)$$

which is the same as a LC model, except for the fact that the class proportions are allowed to differ across higher-level clusters. For $P(\mathbf{y}_{ji}|c)$, as in the LC model, we assume local independence. The higher-level model equals

$$P(\mathbf{y}_j) = \sum_{d=1}^{D} \pi_d \prod_{i=1}^{n_j} P(\mathbf{y}_{ji}|d). \quad (13.16)$$

As can be seen, the main additional model assumptions are that there are D group-level classes and that the individuals' responses within a group are independent given the group's class membership. Combining the above two equations yields the full equation of a two-level LC model:

$$P(\mathbf{y}_j) = \sum_{d=1}^{D} \pi_d \prod_{i=1}^{n_j} \sum_{c=1}^{C} \pi_{c|d} \prod_{k=1}^{K} P(y_{jik}|c).$$

As in a LC class model, we include concomitant variables predicting the higher- and lower-level class memberships, either using a simultaneous analysis or a three-step approach. Moreover, the assumption that the y_{jik} are independent of the groups' class memberships given the individuals' class memberships can be tested and relaxed (Nagelkerke, Oberski, and Vermunt, 2016). Section 13.3.2 presents an application of the multilevel LC model, with two levels.

13.3. APPLICATIONS IN INTERNATIONAL MARKETING

We illustrate the concepts discussed in section 13.2 in the context of international marketing. Ter Hofstede et al. (1999) point out that international marketing has become more important for developing, positioning and selling products. This results from globalization, implying that firms are reacting to international competitors in the local market and competing with local competitors in markets outside their country of origin (Yip, 1995). International marketing strategies require thorough understanding of the various national and cross-national markets (Bijmolt et al., 2004; Ter Hofstede at al., 1999). For example, firms that operate internationally may aim to detect segments that occur across multiple countries, allowing the development of internationalized marketing strategies (Ter Hofstede et al., 1999). Conversely, firms operating in a single country can detect and target segments that only occur in their country and are therefore more likely to react positively to a local strategy implemented by a local company. Detection of such cross-national or country-specific segments relies on segmentation techniques that accommodate for the hierarchy in international data (Bijmolt et al., 2004, Paas et al., 2015; Ter Hofstede et al., 1999). Note that international data are hierarchical in the sense that there are units at two levels, consumers or respondents at the lower level and country represents are the higher level unit (Bijmolt et al., 2004)

We discuss two international segmentation applications of mixture modeling, using a LC model, category (4) in Table 13.1, and the multilevel LC model, (category 5). We show how data from multiple countries can be analyzed in different ways, without ignoring similarities and differences across the countries analyzed and thereby taking into regard the hierarchical nature of international data.

13.3.1. International segmentation using a covariate

The first discussed application (Ter Hofstede et al., 1999) involves a LC model that accommodates the incorporation of within- and cross-country heterogeneity by using country membership of the individual respondent as a concomitant variable or covariate, as discussed in section13.2.6. Ter Hofstede et al. (1999) analyse data on consumer's yoghurt preferences in 11 European Union (EU) countries. Data were part of an EU survey conducted for the European commission; refer to Ter Hofstede et al. (1999) for details. The segmentation basis concerns variables describing consumer Means-End Chains (MECs). MECs assume that consumers obtain desired ends, represented as values through the attributes of specific types of yoghurt. In MECs product attributes are linked to values via benefits (Gutman, 1982). Thus, in this hierarchy there are two key links: attribute-benefit (AB) and benefit-value (BV). As an example of an AB link, Ter Hofstede et al. (1999) find that the product attribute *low fat* is strongly linked to the benefit *good health*. A related BV-link connects *good health* to the value *security*. The AB and BV links vary in strength from 0 to 1. Similarities and differences between consumers, concerning strengths of the links, are used to segment these individuals.

In the resulting LC model (category 4) Ter Hofstede et al. (1999) obtain four segments, using information criteria (see section 13.2.5), with varying strengths in the links. As an example, their segment two is defined by consumers who chose yoghurt mostly accruing to fulfilment of the value *fun and enjoyment*. Through BV-links we find that the value *fun and enjoyment* is linked to the following benefits: *convenient to use, good taste, good quality, good health, good for digestion* and *diet*. In turn the AB-links show that the abovementioned segment 2 benefits are strongly linked to the attributes: *individually packed, with fruit, high priced, mild, organically produced, biobifidus* and *low fat*. Refer to Ter Hofstede et al. (1999) for a more comprehensive discussion of segment two and the other three segments.

Most relevant for the current discussion is that differences in the occurrence of the four segments across the 11 analyzed EU countries are accommodated for using a country-covariate in the LC model (see section 13.2.6.), which leads to different segment sizes across the countries.

Note that Ter Hofstede et al. (1999) included other covariates in their model, which are not discussed in our chapter. Table 13.6 shows segment sizes in the 11 analyzed countries. Large proportions of consumers are in segment 4 across all analyzed countries. Therefore, a cross-national marketing strategy can be developed targeting segment 4 across the 11 EU countries analyzed. Contrarily, the previously discussed segment 2 with the dominant fun and enjoyment value is particularly large in Germany and is also relatively large in Portugal and France. Strategies aimed at targeting this segment are most important for the first mentioned country and can possibly also be applied in France and Portugal. Refer to Ter Hofstede et al. (1999) for a more comprehensive discussion on implications.

*Table 13.6: Occurrence of segments in Ter Hofstede et al study**

Country	p(S1)	p(S2)	p(S3)	p(S4)
Belgium	12.2	17.5	8.5	61.8
Denmark	47.6	2.6	27.0	22.8
France	2.4	21.8	3.3	72.4
Germany	3.2	45.1	26.3	25.4
Great Britain	41.2	7.0	26.0	25.9
Greece	28.2	13.4	6.5	52.0
Ireland	40.6	12.1	17.9	29.5
Italy	5.9	10.2	5.1	78.8
Netherlands	31.5	14.3	17.9	36.3
Portugal	35.2	27.8	4.9	32.1
Spain	24.1	8.5	3.8	62.6

*Note that these figures are copied from Table 5 in Ter Hofstede et al. (1999, p.8)

13.3.2. Multilevel international segmentation

The analysis reported by Ter Hofstede et al. (1999) can be applied to international data with any number of countries. However, the interpretation of the covariate effects becomes cumbersome as the number of higher-level units, such as countries, increases. Under such circumstances, it may be more insightful to simultaneously cluster lower-level units, consumers, and higher-level units, countries, using a multilevel LC model (category 5). Bijmolt et al. (2004) applied this model to consumer ownership data on eight financial products across 15 EU countries. Two countries are split over two regions, i.e., Germany over East and West Germany and Great Britain over Northern

Ireland versus the rest, resulting in 17 higher-level units. Paas et al. (2015) and Paccagnella and Varriale (2013) reported similar, less extensive analyses.

Bijmolt et al. (2004) apply the multilevel LC model to the Eurobarometer 56.0 database, resulting in 14 consumer-level segments and seven higher-level segments for categorizing the 17 countries and regions, using information criteria (see section 13.2.5). That is, models with up to 15 lower level classes, for clustering respondents, and eight higher level classes, for clustering countries, were estimated. Amongst the 120 estimated model the multilevel LC model with 14 respondent segments and seven higher level country-segments led to the lowest value on the information criterion, CAIC. Hence this model has the optimal fit to the data and was selected as the final model. Refer to Bijmolt et al. (2004) for a further description of the analysed data and the model selection.

The 14 consumer-level segments are defined by differing propensities for owning each of the 8 analyzed financial products. For example, in the highly active respondent-level segment 14, we find 100% of the respondents owning the current account, 85.7% a savings account, 83.3% a credit card, 87.1% other bank card, 100% a cheque book, 60.3% an overdraft, 89.3% a mortgage and 39.2% a loan. Contrarily, in segment 1 we find that 4.9% of the respondents owns the current account, 38.7% a savings account, 0.0% a credit card, 0.2% other bank card, 8.5% a cheque book, 0.0% an overdraft, 0.8% a mortgage and 2.2% a loan. Covariate effects on the 14 respondent-level segments are incorporated in the model, as discussed in section 13.2.6. Bijmolt et al. (2004) find that respondents aged 30-59, with above average incomes and with a partner are overrepresented in the consumer segments that are characterised by high penetrations of the products, such as segment 14 that is discussed above.

Next to incorporating covariates, Bijmolt et al (2004) also assessed alternative model specifications, linked to the respondent-level segmentation, such as those discussed in section 13.2.3. Results of previously published studies on developments in consumer financial product portfolios suggest that most consumers will acquire financial products in a similar order, those products relevant for the satisfaction of more basic financial needs before more advanced products

(Dickenson and Kirzner, 1986; Kamakura et al., 1991; Paas and Molenaar, 2005). Bijmolt et al. (2004) test whether such an order of acquisition applies for the eight financial products in their 15-country data set by assessing whether IRT-based assumptions, as those presented in section 13.2.3, lead to better a model, according to the information criteria in section 13.2.5. The model without the restrictions leads to lower information criterion values, CAIC, thus, a common order of acquisition in the Bijmolt et al. (2004) data is not supported.

Bijmolt et al. (2004) also present results of applying a multilevel LC model to internationally collected consumer data. Instead of using covariates to assess similarities and differences in segmentation structures across countries analyzed, the countries themselves are clustered on the basis of these similarities and differences. That is, countries with similar respondent-level segmentation structures are allocated to the same higher-order segment. We present the country clustering reported in Bijmolt et al. (2004) in our Table 13.7. As pointed out by Bijmolt et al. (2004), the classification of countries in Table 13.7 reflects European geography to a certain degree, which supports face validity of the results.

*Table 13.7: Higher level country segments**

Country Segment	Relative Size	Country	Prob.
1	0.256	Belgium, Germany (East and West), The Netherlands Luxembourg	100% 81.1%
2	0.260	Austria, Denmark, Sweden, Finland Luxembourg	100% 18.9%
3	0.175	Great Britain, Ireland, Northern Ireland	100%
4	0.119	Italy, Portugal	100%
5	0.064	Spain	100%
6	0.064	Greece	100%
7	0.064	France	100%

* The content of this Table is derived from Table 3 in Bijmolt et al. (2004, p. 330).

13.4. SOFTWARE

One of the first LC analysis programs, MLLSA, made available by Clifford Clogg in 1977, was limited to a relatively small number of nominal variables. Today's programs can handle many more variables, as well as other scale types. For example, the LEM program (Vermunt, 1997) provides a command language that can be used to specify a large variety of models for categorical data, including LC models. Mplus is a command language based structural equation modeling package that implements many types of LC and mixture models. In addition, routines for the estimation of specific types of LC models are available as SAS, R, and Stata packages/macros (see, for example, Lanza et al., 2007; Skrondal and Rabe-Hesketh, 2004).

Latent GOLD (Vermunt and Magidson, 2000-2016) is a program that was specifically developed for LC analysis, and which contains both an SPSS-like point-and-click user interface and a syntax language. It implements all important types of LC models, such as models for response variables of different scale types, restricted LC models, models with predictors, models with local dependencies, models with multiple discrete latent variables, LC path models, LC Markov models, mixture factor analysis and IRT, and multilevel LC models, as well as features for dealing with partially missing data, performing bootstrapping, performing simulation studies, power computation, and dealing with complex samples.

13.5. SUMMARY AND OTHER APPLICATIONS

We introduced different types of mixture models that can be used for classifying entities, such as consumers. Consecutively we discussed the simple mixture model for a univariate response and extended this to the LC model for multivariate observed responses. Then we discussed the restrictions that may be imposed on the observed responses followed by more technical issues, i.e., parameter estimation and model selection. After this we discussed the LC model for multivariate responses, restricted LC models, LC models with concomitant variables and the multilevel LC model for accommodating data with lower- and higher- level units. The LC model and the multilevel LC model were then illustrated in the context of international segmentation, in which we emphasized the use of a covariate or alternatively a multilevel structure to accommodate for similarities and differences between consumers within and across countries. Last of all software programs applicable for mixture modelling were briefly pointed out.

To conclude this contribution we point out that the marketing literature reports many other applications of the LC model (category 4 in Table 13.1). A selection of these applications is presented in Table 13.8, in order to illustrate some of the possibilities. Note that this list is far from exhaustive. The other category of mixture model discussed in this chapter was the multilevel LC model (category 5 in Table 13.1). We did not find other applications of this model in the marketing literature, besides the application discussed in 13.3.2. However, applications of multilevel LC models to marketing issues have been presented in statistical journals, e.g., Paas et al. (2015) and Paccagnella and Varriale (2013). Thus, this paper may be concluded by pointing out that the marketing research applications of the various types of multilevel LC models should also be exploited.

Table 13.8: Studies based on standard latent class analysis model

Paper	Description
Green, Carmone, Wachspress, JCR, 1976	Modelling consumer adoption of a new telecommunications service
Feick, JMR, 1987	Analysis of behavioral hierarchies in terms of consumer complaining behavior
Grover, Srinivasan, JMR, 1987	Assessing the competitive market structure of different brands in the coffee product category
Kamakura, Novak, JCR, 1992	Identifying segments based on consumer values in the LOV instrument
Kamakura, Wedel, Agrawal, IJRM, 1994	Modeling consumer preferences for bank services
Gupta and Chintagunta, JMR, 1994	Analyzing the relationship between segments characterized by profiles of brand preferences and marketing variable sensitivity in relation to household demographics
Bhatnagar, Ghose, JBR, 2004	Segmentation of web-shoppers based on their purchase behavior across various product categories
Paas and Molenaar, IJRM, 2005	Analysing orders in which consumers acquire financial products – Acquisition Pattern Analysis
Kamakura, Mazzon, IJRM, 2013	Social-economic stratification of consumers to explain consumption of various product categories
De Keyser, Schepers, Komus, IJRM, 2015	Clustering respondents on the basis of self-reported after-sales channel usage

REFERENCES

Bakk, Zs., Tekle, F.B., and Vermunt, J.K. (2013). Estimating the association between latent class membership and external variables using bias adjusted three-step approaches. *Sociological Methodology*, 43, 272-311.

Bhatnagar, A., Ghose, S. (2004). Segmenting consumers based on the benefits and risks of Internet shopping. *Journal of Business Research*, 57 (12), 1352-1360.

Bijmolt, T.H., Paas, L.J., Vermunt, J.K. (2004). Country and consumer segmentation: Multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing*, 21, 323-340.

Bolck, A., M.A. Croon, and J.A. Hagenaars. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, 12, 3-27.

Croon, M. A. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology*, 43, 171-192.

Dayton, C. M. (1999). *Latent class scaling analysis*. Thousand Oaks: Sage Publications.

Dayton, C. M. & Macready, G. B. (1988). Concomitant-variable latent class models*. Journal of the American Statistical Association*, 83, 173-178.

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

Dickenson, J.R. and Kirzner, E. (1986). Priority patterns of acquisition of financial assets. *Journal of the Academy of Marketing Science*, 14 (2), 43-49.

Feick, L.F. (1987). Latent class models for the analysis of behavioral hierarchies. *Journal of Marketing Research*, 24 (May), 174-186.

Green, P.E., Carmone, F.J., Wachspress, D.P. (1976), Consumer segmentation via Latent Class Analysis. *Journal of Consumer Research*, 3 (December), 170-174.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215–231.

Grover, R., Srinivasan, V. (1987). A simultaneous approach to market segmentation and market structuring. *Journal of Marketing Research*, 24 (May), 139-153.

Gupta, S., Chintagunta, P.K. (1994). On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research*, 31 (Feb), 128-136.

Gutman, J. (1982). A means-end model based on consumer categorization processes. *Journal of Marketing*, 46 (Spirng), 60-72.

Hagenaars, J.A. (1988). Latent structure models with direct effects between indicators: local dependence models. *Sociological Methods and Research*, 16, 379-405.

Hambleton, R.K., Swaminathan, H. (1985). *Item Response Theory: Principles and applications*. Thousand Oaks: Sage.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences.* Thousand Oaks, CA: Sage.

Kamakura, W.A., Mazzon, J.A. (2013). Socioeconomic status and consumption in an emerging economy. *International Journal of Research in Marketing*, 30 (1), 4-18.

Kamakura, W.A., Ramaswami, S.N., Srivastava, R.K. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling financial products. *International Journal of Research in Marketing*, 8 (4), 329-349.

Kamakura, W.A., Wedel, M., and Agrawal, J. (1994). Concomitant variable latent class models for the external analysis of choice data. *International Journal of Marketing Research*, 11, 541-464.

Lanza, S.T., Collins, L.M., Lemmon, D.R., and Schafer, J.L. (2007). PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling*, 14, 671 – 694.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis & the interpretation and mathematical foundation of latent structure analysis. In S. A. Stouffer, et al. (Eds.), *Measurement and prediction* (pp. 362–472). Princeton, NJ: Princeton University Press.

Leeflang, P.S.H., Wittink, D.R., Wedel, M., Naert, P.A. (2000). *Building models for marketing decisions*. Dordrecht: Kluwer.

Lindsay, B., Clogg, C. C., and Grego, J. (1991). Semiparametric estimation in the Rasch model and related models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, 86, 96-107.

McLachlan, G.J., and Peel, D. (2000). *Finite mixture models*. New York: John Wiley & Sons.

Magidson, J., and Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology*, 31, 223–264.

Nagelkerke, E., Oberski, D.L., and Vermunt (2016). Goodness-of-fit measures for multilevel latent class models. *Sociological Methodology*.

Oberski, D.L, van Kollenburg, G.H., and Vermunt, J.K. (2013). A Monte Carlo evaluation of three methods to detect local dependence in binary data latent class models. *Advances in Classification and Data Analysis*, 7, 267-279.

Paas, L.J., Bijmolt, T.H.A. and Vermunt, J.K. (2015). Long-term developments of respondent financial product portfolios in the EU: A multilevel latent class analysis. *Metron*, 73 (2). 249-262.

Paas, L.J., Molenaar, I.W. (2005). Analysis of acquisition patterns: A theoretical and empirical evaluation of alternative methods. *International Journal of Research in Marketing*, 22 (1), 87-100.

Paas, L.J., Sijtsma, K., (2008), "Nonparametric Item Response Theory for evaluating the dimensionality of marketing measurement scales: A SERVQUAL application", *Marketing Letters*, 19 (2), 157-170.

Paccagnella, O., Varriale, R. (2013). Asset ownership of elderly across Europe: A multilevel latent class analysis to segment country and respondent. *Advances in Theoretical and Applied Statistics*, 383-393.

Proctor, C. H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika*, 35, 73-78.

Rossiter, J.R., Bellman, S. (2005). *Marketing communications: Theory and applications*. Frenchs Forest: Prentice Hall.

Rost, J. (1990). Rasch models in latent classes. An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.

Sijtsma, K., and Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks: Sage.

Smith, W.R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21 (July), 3-8.

Skrondal, A., and Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. London: Chapman & Hall/CRC.

Ter Hofstede, F., Steenkamp, J.-B.E.M. and Wedel, M. (1999). International segmentation based on consumer-product relations. *Journal of Marketing Research*, 36 (February), 1-17.

Vermunt, J.K. (2001). The use restricted latent class models for defining and testing nonparametric and parametric IRT models. *Applied Psychological Measurement*, 25, 283-294.

Vermunt, J.K. (2003). Multilevel latent class models. *Sociological Methodology*, 33, 213-239.

Vermunt, J.K., and Magidson, J. (2002). Latent class cluster analysis. In: J.Hagenaars and A.McCutcheon (eds.). *Applied latent class analysis*, 89-106. Cambridge University Press.

Vermunt, J.K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18, 450-469.

Vermunt, J.K., and Magidson, J. (2000-2016). *Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax*. Belmont Massachussetts: Statistical Innovations Inc.

Wedel, M. and Kamakura, W.A. (2000). Market segmentation: Conceptual and methodological foundations, 2nd Edition. Dordrecht: Kluwer.

Yip, G.S. (1995). *Total global strategy*. Englewood Cliffs NJ: Prentice Hall.