Multilevel Latent Class Models

Multilevel Latent Class Models

Jeroen K. Vermunt

Department of Methodology and Statistics, Tilburg University

Address of correspondence:

Jeroen K. Vermunt

Department of Methodology and Statistics

Tilburg University

P.O.Box 90153

5000 LE Tilburg

The Netherlands

E-mail: J.K.Vermunt@uvt.nl

Accepted by Sociological Methodology (November 4, 2002)

Multilevel Latent Class Models

Abstract

Latent class (LC) models developed so far assume that observations are independent. Parametric and nonparametric random-coefficient LC models are proposed that make it possible to relax this assumption. The models can, for example, be used for the analysis of data collected with complex sampling designs, data with a multilevel structure, and multiple-group data for more than a few groups. An adapted EM algorithm is presented that makes maximum likelihood estimation feasible. The new model is illustrated with examples from organizational, educational, and cross-national comparative research.

Multilevel Latent Class Models

1 Introduction

In the past decade, latent class (LC) analysis (Lazarsfeld, 1950; Goodman, 1974) has become a more widely used technique in social sciences research. One of its applications is clustering or constructing typologies with observed categorical variables. Another related application is dealing with measurement error in nominal and ordinal indicators. An important limitation of the LC models developed so far is, however, that they assume that observations are independent, an assumption that is often violated. This paper presents a random-coefficients or multilevel LC model that makes it possible to relax this assumption.

Random-coefficients models can be used to deal with various types of dependent observations (Agresti et al., 2000). An example is dependent observations in data sets collected by two-stage cluster sampling or longitudinal designs. A more theory based use of random-coefficients models that is popular in fields such as educational and organizational research is often referred to as multilevel or hierarchical modeling, a method that is intended to disentangle group-level from individual-level effects (Bryk and Raudenbush, 1992; Goldstein, 1995; Snijders and Bosker, 1999). Another interesting application of these methods is in the context of multiple-group analysis, such as in cross-national comparative research based on data from a large number of countries (see, for example, Wong and Mason, 1985).

In a standard LC model, it is assumed that the model parameters are the same for all persons (level-1 units). The basic idea of a multilevel LC model is that some of the model parameters are allowed to differs across groups, clusters, or level-2 units. For example, the probability of belonging to a certain latent class may differ across organizations or countries. Such differences can be modelled by including group dummies in the model, as is done in multiple-group LC analysis (Clogg and Goodman, 1984), which amounts to using what is called a fixed-effects approach. Alternatively, in a random-effects approach, the group-specific coefficients are assumed to come from a particular distribution, whose parameters should be estimated. Depending on whether the form of the mixing distribution is specified or not, either a parametric or a nonparametric random-effects approach is obtained.

The proposed multilevel LC model is similar to a random-coefficients logistic regression model (Wong and Mason, 1985; Hedeker and Gibbons, 1996; Hedeker, 1999; Agresti et al., 2000). A difference is that the dependent variable is not directly observed, but a latent variable with several observed indicators. The model can, therefore, be seen as an extension of a random-coefficients logistic regression model in which there is measurement error in the dependent variable. It is well-known that LC models can be used to combine the information contained in multiple outcome variables (Bandeen-Roche et al., 1997).

The model is also similar to the multilevel item response theory (IRT) model that was recently proposed by Fox and Glas (2001). A conceptual difference is, however, that in IRT models the underlying latent variables are assumed to be continuous instead of discrete. Because of the similarity between LC and IRT models (see, for example, Heinen, 1996), it is not surprising that restricted multilevel LC models can be used to approximate multilevel IRT models.

The idea of introducing random effects in LC analysis is not new: see, for example, Qu, Tan, and Kutner (1996), and Lenk and DeSarbo (2000). The models proposed by these authors are, however, not multilevel models, and therefore conceptually and mathematically very different from the models described in this paper. Although there is no option for combining LC structures with random effects in the current version of the GLLAMM program of Rabe-Hesketh, Pickles, and Skrondal (2001), the model I propose fits very naturally into the general multilevel modeling framework developed by these authors.

The multilevel LC model can be represented as a graphical or path model containing one latent variable per random coefficient and one latent variable per level-1 unit within a level-2 unit. The fact that the model contains so many latent variables makes the use of a standard EM algorithm for maximum likelihood (ML) estimation impractical. The ML estimation problem can, however, be solved by making use of the conditional independence assumptions implied by the graphical model. More precisely, I adapted the E step of the EM algorithm to the structure of the multilevel LC model.

The next section describes the multilevel LC model. Then, attention is paid to estimation issues that are specific for this new model. Section 4 presents applications from organizational, educational, and cross-national comparative research. The paper ends with a short discussion.

2 The multilevel LC model

Let Y_{ijk} denote the response of individual or level-1 unit i within group or level-2 unit j on indicator or item k. The number of level-2 units is denoted by J, the number of level-1 units within level-2 unit j by n_j, and the number of items by K. A particular level of item k is denoted by s_k and its number of categories by S_k. The latent class variable is denoted by X_{ij}, a particular latent class by t, and the number of latent classes by T. Notation $\mathbf{Y}_{ij}$ is used to refer to the full vector of responses of case i in group j, and $\mathbf{s}$ to refer to a possible answer pattern.

The probability structure defining a simple LC model can be written down as follows:

$$\begin{aligned} P(\mathbf{Y}_{ij} = \mathbf{s}) &= \sum_{t=1}^{T} P(X_{ij} = t) P(\mathbf{Y}_{ij} = \mathbf{s} | X_{ij} = t) \\ &= \sum_{t=1}^{T} P(X_{ij} = t) \prod_{k=1}^{K} P(Y_{ijk} = s_k | X_{ij} = t). \end{aligned} \tag{1}$$

The probability of observing a particular response pattern, $P(\mathbf{Y}_{ij} = \mathbf{s})$, is a weighted average of class-specific probabilities $P(\mathbf{Y}_{ij} = \mathbf{s}|X_{ij} = t)$. The weight $P(X_{ij} = t)$ is the probability that person i in group j belongs to latent class t. As can be seen from the second line, the indicators Y_{ijk} are assumed to be independent of each other given class membership, which is often referred to as the local independence assumption. The term $P(Y_{ijk} = s_k|X_{ij} = t)$ is the probability of observing response s_k on item k given that the person concerned belongs to latent class t. These conditional response probabilities are used to name the latent classes.

The general definition in equation (1) applies to both the standard and the multilevel LC model. In order to be able to distinguish the two, the model probabilities have to be written in the form of logit equations. In the standard LC model,

$$P(X_{ij} = t) = \frac{\exp(\gamma_t)}{\sum_{r=1}^{T} \exp(\gamma_r)} \tag{2}$$

$$P(Y_{ijk} = s_k|X_{ij} = t) = \frac{\exp(\beta^k_{s_k t})}{\sum_{r=1}^{S_k} \exp(\beta^k_{rt})}. \tag{3}$$

As always, identifying constraints have to be imposed on the logit parameters, for example, $\gamma_1 = \beta^k_{1t} = 0$.

The fact that the γ and β parameters appearing in equations (2) and (3) do not have an index j indicates that their values are assumed to be independent of the group to which one belongs. Taking into account the multilevel structure involves relaxing this assumption. The most general multilevel LC model is obtained by assuming that all model parameters are group specific; that is,

$$P(X_{ij} = t) = \frac{\exp(\gamma_{tj})}{\sum_{r=1}^{T} \exp(\gamma_{rj})} \tag{4}$$

$$P(Y_{ijk} = s_k|X_{ij} = t) = \frac{\exp(\beta^k_{s_k tj})}{\sum_{r=1}^{S_k} \exp(\beta^k_{rtj})}. \tag{5}$$

Without further restrictions, this model is equivalent to an unrestricted multiple-group LC model (Clogg and Goodman, 1984). A more restricted multiple-group LC model is obtained by assuming that the item conditional probabilities do not depend on the level-2 unit; that is, by combining specifications (4) and (3). In practice, such a partially heterogeneous model assuming invariant measurement error is the most useful specification, although it is not a problem to relax this assumption for some of the indicators.

It will, however, be clear that such a multiple-group or fixed-effects approach may be problematic if there are more than a few groups because group-specific estimates have to be obtained for certain model parameters. Not only the number of parameters to be estimated increases rapidly with the number of level-2 units, the estimates may also be very unstable with group sizes that are typical in multilevel research. Another disadvantage of the fixed-effects approach is that all group differences are "explained" by the group dummies, making it impossible to determine the effects of level-2 covariates on the probability of belonging to a certain latent class. Below I show how to include such covariates in the model.

A parametric approach

The problems associated with the multiple-group approach can be tackled by adopting a random-effects approach: rather than estimating a separate set of parameters for each group, the group-specific effects are assumed to come from a certain distribution. Let us look at the simplest case: a two-class model with group-specific class-membership probabilities as defined by equation (4), and with $\gamma_{1j} = 0$ for identification. Typically, random coefficients are assumed to come from a normal distribution, yielding a LC model in which

$$\gamma_{2j} = \gamma_2 + \tau_2 \cdot u_j, \tag{6}$$

with $u_j \sim N(0, 1)$. Note that this amounts to assuming that the between-group variation in the log odds of belonging to the second instead of the first latent class follows a normal distribution with a mean equal to γ_2 and a standard deviation equal to τ_2.

With more than two latent classes, one has to specify the distribution of the $T-1$ random-intercept terms γ_{tj}. One possibility is to work with a $(T-1)$-dimensional normal distribution. Another option is to use more a restricted structure of the form,

$$\gamma_{tj} = \gamma_t + \tau_t \cdot u_j, \tag{7}$$

with $u_j \sim N(0,1)$ and with one identifying constraint on the γ_t and another on the τ_t: for example, $\gamma_1 = \tau_1 = 0$. This specification was also used by Hedeker (1999) in the context of random-effects multinomial logistic regression analysis. The implicit assumption that is made is that the random components in the various γ_{tj} are perfectly correlated. More specifically, the same random effect u_j is scaled in a different manner for each t by the unknown τ_t. This formulation, which is equivalent to Bock's nominal response model (Bock, 1972), is based on the assumption that each nominal category is related to an underlying latent response tendency. Whether such a restricted structure suffices or a more general formulation is needed is an empirical issue. Note that (7) reduces to (6) if $T = 2$.

A measure that often of interest in random-effects models is the intraclass correlation. It is defined as the proportion of the total variance accounted for by the level-2 units, where the total variance equals the sum of the level-1 and level-2 variances. Hedeker (in press) showed how to compute the intraclass correlation in random-coefficients multinomial logistic regression models. The same method can be used in multilevel LC models; that is,

$$r_{It} = \frac{\tau_t^2}{\tau_t^2 + \pi^2/3}. \tag{8}$$

This formula makes use of the fact that the level-1 variance can be set equal to the variance of the logistic distribution, which equals $\pi^2/3 \approx 3.29$. Notice that $T-1$ independent intraclass correlations can be computed.

A nonparametric approach

A disadvantage of the presented random-effects approach is that it makes quite strong assumptions about the mixing distribution. An attractive alternative is, therefore, to work with a discrete unspecified mixing distribution. This yields a nonparametric random-coefficients LC model in which there are not only latent classes of level-1 units but also latent classes of level-2 units sharing the same parameter values. Such an approach does not only have the advantage of less strong distributional assumptions and less computational burden (Vermunt and Van Dijk, 2001), it may also fit better to the substantive research problem at hand. In many settings, it is more natural to classify groups (for example, countries) into a small number of types than to place them on a continuous scale.

It should be noted that with nonparametric I do not mean "distribution free". In fact, the normal distribution assumption is replaced by a multinomial distribution assumption. According to Laird (1978), a nonparametric characterization of the mixing distribution is obtained by increasing the number of mass points till a saturation point is reached. In practice, however, one will work with fewer latent classes that the maximum number that can be identified.

Let W_j denote the value of group j on the latent class variable defining the discrete mixing distribution. In a nonparametric approach, the model for the latent class probability equals

$$P(X_{ij} = t | W_j = m) = \frac{\exp(\gamma_{tm})}{\sum_{r=1}^{T} \exp(\gamma_{rm})}, \tag{9}$$

where m denotes a particular mixture component. Besides the M component-specific coefficients, we have to estimate the size of each component, denoted by π_m. Note that we can write γ_{tm} as

$$\gamma_{tm} = \gamma_t + u_{tm}, \tag{10}$$

where the u_{tm} come from an unspecified distribution with M mass points.

Covariates

A natural extension of the random-coefficient LC model involves including level-1 and level-2 covariates to predict class membership. Suppose that there is one level-2 covariate Z_{1j} and one level-1 covariate Z_{2ij}. A multinomial logistic regression model for X_{ij} with a random intercept is obtained by

$$P(X_{ij}=t|Z_{1j},Z_{2ij})=\frac{\exp(\gamma_{0tj}+\gamma_{1t}Z_{1j}+\gamma_{2t}Z_{2ij})}{\sum_{r=1}^{T}\exp(\gamma_{0rj}+\gamma_{1r}Z_{1j}+\gamma_{2r}Z_{2ij})}.$$

This model is an extension of the LC model with concomitant variables proposed by Dayton and McReady (1988); that is, a model containing not only fixed but also random effects.

Not only the intercept, but also the effects of the level-1 covariates may be assumed to be random coefficients. A model with a random slope is obtained by replacing γ_{2t} with γ_{2tj}, and making certain distributional assumptions about γ_{2tj}. In fact, any multilevel model that can be specified for an observed nominal outcome variable can also be applied with the latent class variable, which is in fact an indirectly observed nominal outcome variable.

Also the nonparametric approach can easily be extended to include level-1 and level-2 covariates. An example is

$$P(X_{ij}=t|Z_{1j},Z_{2ij},W_j=m)=\frac{\exp(\gamma_{0tm}+\gamma_{1tm}Z_{1j}+\gamma_{2t}Z_{2ij})}{\sum_{r=1}^{T}\exp(\gamma_{0rm}+\gamma_{1rm}Z_{1j}+\gamma_{2r}Z_{2ij})}.$$

In this model, both the intercept and the slope of the level-1 covariate are assumed to depend on the mixture variable W_j.

Item bias

The last extension I would like to mention is the possibility to allow for group differences in the class-specific conditional response probabilities, as was already indicated in equation (5). It may happen that certain items are responded in a different manner by individuals belonging to different groups, a phenomenon that is sometimes referred to as item bias.

A random-effects specification for the item bias in item k can, for example, be of the form

$$\beta^k_{s_k tj} = \beta^k_{s_k t} + \sigma_s \cdot e_j,$$

where the σ_s are category-specific parameters, and $e_j \sim N(0,1)$. One identifying constraint has to be imposed on the $\beta^k_{s_k t}$ and on the σ_t parameters: for example, $\beta^k_{1t} = \sigma_1 = 0$. Note that this specification is similar to the one that was proposed for γ_{tj} in equation (7). The terms u_j and e_j can be assumed to be correlated or uncorrelated.

In the nonparametric approach, item bias can be dealt with by allowing the conditional response probabilities to depend on the mixture variable; i.e.,

$$P(Y_{ijk} = s_k | X_{ij} = t, W_j = m) = \frac{\exp(\beta^k_{s_k tm})}{\sum_{r=1}^{S_k} \exp(\beta^k_{rtm})}.$$

In such a specification, latent classes of groups not only differ with respect to the latent class distribution of individuals, but also with respect to item distributions within latent classes of individuals.

3 Maximum likelihood estimation

The parameters of the multilevel LC model can be estimated by maximum likelihood (ML). The likelihood function is based on the probability density for the data of a complete level-2 unit, denoted by $P(\mathbf{Y}_j|\mathbf{Z}_j)$. It should be noted that these are independent observations while observations within a level-2 unit are not assumed to be independent. The log-likelihood to be maximized equals

$$\log L = \sum_{j=1}^{J} \log P(\mathbf{Y}_j|\mathbf{Z}_j).$$

In a parametric random-coefficient LC model, the relevant probability density equals

$$\begin{aligned} P(\mathbf{Y}_j|\mathbf{Z}_j) &= \int_{\mathbf{u}} P(\mathbf{Y}_j|\mathbf{Z}_j, \mathbf{u}, \boldsymbol{\theta}) f(\mathbf{u}|\boldsymbol{\theta}) d\mathbf{u} \\ &= \int_{\mathbf{u}} \left\{ \prod_{i=1}^{n_j} P(\mathbf{Y}_{ij}|\mathbf{Z}_{ij}, \mathbf{u}, \boldsymbol{\theta}) \right\} f(\mathbf{u}|\boldsymbol{\theta}) d\mathbf{u}, \qquad (11) \end{aligned}$$

where $\boldsymbol{\theta}$ denotes the complete set of unknown parameters to be estimated and $f(\mathbf{u}|\boldsymbol{\theta})$ the multivariate normal mixing distribution. As can be seen, the responses of the n_j level-1 units with level-2 unit j are assumed to be independent of one another given the random coefficients. Of course, the contributions of the level-1 units have the form of a LC model; that is,

$$\begin{aligned} P(\mathbf{Y}_{ij}|\mathbf{Z}_{ij},\mathbf{u},\boldsymbol{\theta}) &= \sum_{t=1}^{T} P(X_{ij}=t,\mathbf{Y}_{ij}|\mathbf{Z}_{ij},\mathbf{u},\boldsymbol{\theta}) \\ &= \sum_{t=1}^{T} P(X_{ij}=t|\mathbf{Z}_{ij},\mathbf{u},\boldsymbol{\theta}) \prod_{k=1}^{K} P(Y_{ijk}|X_{ij}=t,\mathbf{Z}_{ij},\mathbf{u},\boldsymbol{\theta}) \end{aligned}$$

The integral at the right-hand side of equation (11) can be evaluated by the Gauss-Hermite quadrature numerical integration method (Stroud and Secrest, 1966; Bock and Aitkin, 1981). After orthonormalizing the random coefficients, the multivariate normal mixing distribution is approximated by a limited number of M discrete points. Hedeker (1999) provided a detailed description of the implementation of this method in the estimation of random-coefficients multinomial logistic regression models. An equivalent approach is used here. The basic idea is that the integral is replaced by a summation over M quadrature points,

$$\begin{aligned} P(\mathbf{Y}_j|\mathbf{Z}_j) &= \sum_{m=1}^{M} P(\mathbf{Y}_j|\mathbf{Z}_j,\mathbf{u}_m,\boldsymbol{\theta})\pi_m \\ &= \sum_{m=1}^{M} \left\{ \prod_{i=1}^{n_j} P(\mathbf{Y}_{ij}|\mathbf{Z}_{ij},\mathbf{u}_m,\boldsymbol{\theta}) \right\} \pi_m. \end{aligned} \qquad (12)$$

Here, $\mathbf{u}_m$ and π_m denote the fixed optimal quadrature nodes and weights corresponding to the (multivariate) normal density of interest. These can be obtained from published tables (see, for example, Stroud and Secrest, 1966). The integral can be approximated to any practical degree of accuracy by setting M sufficiently large.

As was explained above, it is also possible to use an unspecified discrete mixing distribution. The probability density corresponding to a nonparametric model with M components is similar to the approximate density defined in equation (12). A difference is, however, that now $\mathbf{u}_m$ and π_m are unknown parameters to be estimated, and that the parameter vector $\boldsymbol{\theta}$ no longer contains the variances and covariances of the random coefficients. Equation (10) defines $\mathbf{u}_m$ for the nonparametric case.

Implementation of the EM algorithm

ML estimates of the model parameters can be obtained by an EM algorithm with an E step that is especially adapted to the problem at hand. A standard implementation of the E step would involve computing the joint conditional expectation of n_j latent class variables and the latent variables representing the random effects; that is, a posterior distribution with $M \cdot T^{n_j}$ entries $P(W_j = m, \mathbf{X}_j = \mathbf{t}|\mathbf{Y}_j, \mathbf{Z}_j)$. It will be clear that this is not practical when there are more than a few level-1 units per level-2 unit.

[Insert figure 1 about here]

Figure 1 depicts the graphical model underlying the multilevel LC model. The level-2 latent variable W_j influences the n_j level-1 latent variables X_{ij}, which are mutually independent given W_j. The number of level-1 latent variables differs per group. More precisely, there is one X_{ij} for each each i within group j, which means n_j in total. Each level-1 unit has it own set of indicators Y_{ijk}. The graphical representation shows that, on the one hand, we are dealing with a model with many latent variables, but that, on the other hand, it is a quite restricted model.

The algorithm I developed makes use of the fact that the only thing we need in the E step are the marginal posterior probabilities $P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j)$. This follows from the fact that these are the terms appearing in the complete data log-likelihood. Rather than first computing $P(W_j = m, \mathbf{X}_j = \mathbf{t}|\mathbf{Y}_j, \mathbf{Z}_j)$ and subsequently collapsing over all $X_{i'j}$, $i' \neq i$, in order to obtain $P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j)$, $P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j)$ is computed directly. The procedure makes use of the conditional independence assumptions underlying the graphical model associated with the density defined in equation (12). This yields an algorithm in which computation time increases linearly rather than exponentially with n_j. The proposed method is similar to the forward-backward algorithm for the estimation of hidden Markov models for large numbers of time points (Baum et al., 1970; Juang and Rabiner, 1991). As will become clear below, the method I propose could be called an upward-downward algorithm. Latent variables are integrated or summed out by going from the lower- to the higher-level units, and the marginal posteriors are subsequently obtained by going from the higher- to the lower-level units.

The marginal posterior probabilities $P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j)$ needed to perform the E step can be obtained using the following simple decomposition:

$$P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j) = P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j)P(X_{ij} = t|W_j = m, \mathbf{Y}_j, \mathbf{Z}_j)$$

The proposed efficient algorithm makes use of the fact that in the multilevel LC model

$$P(X_{ij} = t|W_j = m, \mathbf{Y}_j, \mathbf{Z}_j) = P(X_{ij} = t|W_j = m, \mathbf{Y}_{ij}, \mathbf{Z}_{ij}),$$

i.e., given W_j, X_{ij} is independent of the observed (and latent) variables of the other level-1 units within the same level-2 unit. This is the result of the fact that level-1 observations are mutually independent given the random coefficients (given the value of W_j), as is expressed in the density function described in equation (11). This can also be seen from Figure 1, which depicts the conditional independence assumptions underlying the multilevel LC model. Using this important result, we get the following slightly simplified decomposition:

$$P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j) = P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j)P(X_{ij} = t|W_j = m, \mathbf{Y}_{ij}, \mathbf{Z}_{ij}). \quad (13)$$

This shows that the problem of obtaining the marginal posterior probabilities reduces to the computation of the two terms at the right-hand side of this equation. Using the short-hand notation $\pi_{ij|m}$ for $P(\mathbf{Y}_{ij}|\mathbf{Z}_{ij}, \mathbf{u}_m, \boldsymbol{\theta})$ and $\pi_{ijt|m}$ for $P(X_{ij} = t, \mathbf{Y}_{ij}|\mathbf{Z}_{ij}, \mathbf{u}_m, \boldsymbol{\theta})$, and noting that

$\pi_{ij|m} = \sum_{t=1}^{T} \pi_{ijt|m}$, we obtain

$$\begin{aligned} P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j) &= \frac{\pi_m \prod_{i=1}^{n_j} \pi_{ij|m}}{\sum_{m=1}^{M} \pi_m \prod_{i=1}^{n_j} \pi_{ij|m}} \\ P(X_{ij} = t|W_j = m, \mathbf{Y}_{ij}, \mathbf{Z}_{ij}) &= \frac{\pi_{ijt|m}}{\sum_{t=1}^{T} \pi_{ijt|m}} = \frac{\pi_{ijt|m}}{\pi_{ij|m}}. \end{aligned} \quad (14)$$

As can be seen, the basic operation that has to be performed is the computation of $\pi_{ijt|m}$ for each i, j, t, and m. In the upward step, the n_j sets of $\pi_{ijt|m}$ terms are used to obtain $P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j)$. The downward step involves the computation of $P(X_{ij} = t|W_j = m, \mathbf{Y}_{ij}, \mathbf{Z}_{ij})$, as well as $P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j)$ via equation (13). In the upward-downward algorithm computation time increases linearly with the number of level-1 observations instead of exponentially, as would be the case in a standard E step. Computation time can be decreased somewhat more by grouping records with the same values for the observed variables within level-2 units; that is, records with the same value for $\pi_{ijt|m}$.

The remaining part of the implementation of the EM algorithm is as usual. In the E step, the $P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j)$ and $P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j)$ are used to obtain improved estimates for the observed cell entries in the relevant marginal tables. Suppose that we have a multilevel LC model in which the latent class probabilities contain random coefficients and are influenced by covariates $\mathbf{Z}$. Denoting a particular covariate pattern by r, the relevant marginal tables have entries f_m^W, $f_{mtr}^{WX\mathbf{Z}}$, and $f_{ts_k}^{XY_k}$, which can be obtain as follows:

$$\begin{aligned} f_m^W &= \sum_{j=1}^{J} P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j) \\ f_{mtr}^{WX\mathbf{Z}} &= \sum_{j=1}^{J} \sum_{i=1}^{n_j} P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j) I(\mathbf{Z}_j = r) \\ f_{ts_k}^{XY_k} &= \sum_{m=1}^{M} \sum_{j=1}^{J} \sum_{i=1}^{n_j} P(W_j = m, X_{ij} = t|\mathbf{Y}_j, \mathbf{Z}_j) I(Y_{ijk} = s_k); \end{aligned}$$

where $I(Y_{ijk} = s_k)$ and $I(\mathbf{Z}_j = r)$ are 1 if the corresponding condition is fulfilled, and 0 otherwise.

In the M step, standard complete data methods for the estimation of logistic regression models can be used to update the parameter estimates using the completed data as if it were observed.

A practical problem in the above implementation of the ML estimation of the multilevel LC model is that underflows may occur in the computation of $P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j)$. More precisely, because it may involve multiplication of a large number $(1 + n_j \cdot K)$ of probabilities, the numerator of equation (14) may become equal to zero for each m. Such underflows can, however, easily be prevented by working on a log scale. Letting $a_{jm} = \log(\pi_m) + \sum_i^{n_j} \log(\pi_{ij|m})$ and $b_j = \max(a_{jm})$, $P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j)$ can be obtained by

$$P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j) = \frac{\exp\left[a_{jm} - b_j\right]}{\sum_r^M \exp\left[a_{jr} - b_j)\right]}.$$

Posterior means

One of the objectives of a multilevel analysis may be to obtain estimates of the group-specific parameters. Suppose we are interested in the value of γ_{tj} for each level-2 unit. A simple estimator is the posterior mean, denoted by $\overline{\gamma}_{tj}$. It is computed as follows:

$$\begin{aligned}\overline{\gamma}_{tj} &= \gamma_t + \tau_t \cdot \overline{u}_j \\ &= \gamma_t + \tau_t \cdot \sum_{m=1}^{M} u_m P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j). \end{aligned} \qquad (15)$$

Here, u_m denotes the value of the mth quadrature node and $P(W_j = m|\mathbf{Y}_j)$ the posterior probability associated with this node. The quantity $\overline{\gamma}_{tj}$ can subsequently be used to calculate the group-specific latent class distributions by means of equation (4).

In the nonparametric approach, $\overline{\gamma}_{tj}$ is defined as

$$\overline{\gamma}_{tj} = \sum_{m=1}^{M} \gamma_{tm} P(W_j = m|\mathbf{Y}_j, \mathbf{Z}_j); \qquad (16)$$

that is, as a weighted average of the parameters of the M mixture components.

Standard errors

Contrary to Newton-like methods, the EM algorithm does not provide standard errors of the model parameters as a by-product. Estimated asymptotic standard errors can be obtained by computing the observed information matrix, the matrix of second-order derivatives of the log-likelihood function towards all model parameters. The inverse of this matrix is the estimated variance-covariance matrix. For the examples presented in the next section, I computed the necessary derivatives numerically.

The information matrix can also be used to check identifiability. A sufficient condition for local identification is that all the eigenvalues of the information matrix are larger than zero.

Software implementation

The multilevel LC model cannot be estimated with standard software for LC analysis. The upward-downward algorithm described in this section was implemented in an experimental version of the Latent GOLD program (Vermunt and Magidson, 2000). The method will become available in a next version of this program for LC analysis.

4 Three applications

Three applications of the proposed new method are presented. These not only illustrate three interesting application fields, but also the most important model specification options. In the first example, I use data from a Dutch survey in which employees of various teams are asked about their work conditions. Multilevel LC analysis is used to construct a task-variety scale and to determine the between-team heterogeneity of the latent class probabilities. The second example uses a Dutch data set containing information on the mathematical skills of grade 8 pupils from various schools. The research question of interest is as to whether school differences remain after controlling for individual characteristics such as non-verbal intelligence and socioeconomic status of the family. In the third example, I use data from the 1999 European Values Survey. Country differences in the proportion of post-materialists are modelled by means random effects. Contrary to the previous applications, I am not only interested in the overall between-group differences, but also in the latent distribution for each of the groups (countries).

4.1 Organizational research

In a Dutch study on the effect of autonomous teams on individual work conditions, data were collected from 41 teams of two organizations, a nursing home and a domiciliary care organization. These teams contained 886 employees. For the example, I took five dichotomized items of a scale measuring perceived task variety (Van Mierlo et al., 2002). The item wording is as follows (translated from Dutch):

1. Do you always do the same things in your work?
2. Does your work require creativity?
3. Is your work diverse?
4. Does your work make enough usage of you skills and capacities?
5. Is there enough variation in your work?

The original items contained four answer categories. In order simplify the analysis, I collapsed the first two and the last two categories. Because some respondents had missing values on one or more of the indicators, I adapted the ML estimation procedure to deal with such partially observed indicators.

I will analyze this data set by means of LC analysis. This means that I am assuming that the researcher is interested in building a typology of employees based on their perceived task variety. On other hand, if one would be interested in constructing a continuous scale, a latent trait analysis would be more appropriate. Of course, also in that situation the multilevel structure should be taken into account.

In the analysis of this data set, I used a simple unrestricted LC model combined with three types of specifications for the between-team variation in the class-membership probabilities: no random effects, parametric random effects as defined in equation (6), and nonparametric random effects as defined in equation (9).

[Insert Table 1 about here]

Table 1 reports the log-likelihood (LL) value, the number of parameters, and the BIC value for the models that were estimated. I first estimated models without random effects. The BIC values for the one to three class model (Models 1-3) without random effects show that a solution with two classes suffices. Subsequently, I introduced random effects in the two-class model (Models 4-6). From the results obtained with Models 4 and 5, it can be seen that there is clear evidence for between-team variation in the latent distribution: These models have much lower BIC values than the two-class model without random effects. The two-class finite-mixture representation (Model 5) yields a slightly lower LL than the parametric representation of the between-team variation (Model 4), but a somewhat higher BIC because it uses one additional parameter. The three-class finite-mixture model (Model 6) has almost the same LL value as Model 5, which indicates that no more than two latent classes of teams can be identified.

[Insert Table 2 about here]

Table 2 reports the parameter estimates obtained with Models 4 and 5, where the conditional response probabilities describing the relationship between the latent variable and the indicators correspond to the high task-variety response (disagree for item 1 and agree for the others). The first class has a much lower conditional response probability than class two on each of the indicators. The two classes can therefore be named “low task-variety” and “high task-variety”. Note that the conditional response probabilities take on almost the same values in both random-effects models, which shows that the definition of the classes is quite robust for the specification of the mixture distribution.

In the parametric specification of the between-team heterogeneity (Model 4), the mean and standard deviation of the log odds of belonging to class 2 equal 0.73 and 0.87, respectively. In order to get an impression of the meaning of these numbers on the probability scale, one can fill in a value for u_j in equation (6) and substitute the obtained value for γ_{2j} in equation (4). For example, with u_j equal to -1.28 and 1.28, the z values corresponding to the lower and upper 10% tails of the normal distribution, we get latent class probabilities of .41 and .86. These numbers indicate that there is a quite large team effect on the perceived task variety. The intraclass correlation obtained with equation (8) equals .19, which means that 19% of the total variance is explained by team membership.

The mixture components in the two-class finite-mixture model (Model 5) contained 37 and 63 percent of the teams. Their log odds of belonging to the high task-variety class are -.35 and 1.29, respectively. These log odds correspond to probabilities of .41 and .78. Although the parametric and non-parametric approach capture the variation across teams in a somewhat different manner, both show that there are large between-team differences in the individual perception of the variety of the work. The substantive conclusion based on Model 5 would be that there are two types of employees and two types of teams. The two types of teams differ with respect to the distribution of the team members over the two types of employees.

4.2 Educational research

For the second example, I used data from a Dutch study aimed to compare the mathematical skills of grade 8 students across schools (Doolaard, 1999). A mathematical test consisting of

18 items was given to 2156 students of a sample of 97 schools. The 18 dichotomous (correct/incorrect) items are used to construct a LC model measuring individual skills. There is also information on the individual-level covariates socioeconomic status (Ses; standardized), non-verbal intelligence (Isis; standardized), and Gender (0=male, 1=female), and a school-level covariate indicating whether a school participates in the national school leaving examination (Cito; 0=no, 1=yes).

Again I assume that the researcher is interest in constructing a latent typology, in this case of types of students who differ in their mathematical skills. It can be expected that schools differ with respect to the mathematical ability of their students. However, in the Dutch educational context in which schools are assumed to be more or less of the same quality, one would expect that the school effect decreases or may be even disappears if one controls for the composition of the groups. The availability of individual level covariates that can be expected to be strongly related to mathematical ability makes it possible to control for such composition effects.

[Insert Table 3 about here]

Table 3 reports the test results obtained with this data set. I first estimated LC models without covariates nor random effects. The BIC values of these models (Models 1-4) show clearly that more than two latent classes are needed, which is not surprising given the large number of indicators. Because the three-class class model has the lowest BIC, I decided to continue with this solution when introducing covariates and random-coefficients. In the parametric random-effects approach, the random intercepts were modelled as in equation (7). The between-school variation in the latent distribution is captured equally well by a parametric specification and a nonparametric specification with three latent classes (see LL of Models 6-9). As can be seen from the BIC values of Models 10 and 11, introduction of covariates yields a large improvement of the fit compared to Models 6 and 8. Comparison with Model 5, which contains the four covariates but no random effects, shows that the between-school differences remain after controlling for the three level-1 and the single level-2 covariate.

[Insert Table 4 about here]

Table 4 reports the estimated parameters of the multinomial logistic regression model for the latent class variable according to Model 10. These are effects on the log odds of belonging to class 2 or 3 rather than to class 1. The item parameters, which are not reported here, show that the three ordered latent classes can be labelled as "low", "medium", and "high".

Ses, Isis, and Cito have positive effects of both log odds, and Gender has negative effects. This shows that having a higher socioeconomic status, having a higher score on the non-verbal intelligence test, being a male, and belonging to a Cito school increases the probability of belonging to the middle and high ability groups. It can also be seen that there remains a large amount of between-school variation after controlling for the four covariates. The size of the school effect can directly be compared with the effects of the Ses and Isis covariates because these are also standardized variables. The standard deviation of the Gender and Cito dummies are about twice as small, which means that we have to divide their effects by two if we want to compare them with the other ones. The school effect turns out to be as important as the effect of a pupil's non-verbal intelligence (Isis) and more important than the other covariates. The differences between schools can not be explained by differences in the composition of their populations.

4.3 Comparative research

The third example makes use of data from the 1999 European Values Study (EVS) containing information from 32 countries. I took a sample of 10% of the available cases per countries, yielding 3584 valid cases. An popular scale in the survey is the post-materialism scale developed by Inglehart (1977). An important issue has always been the comparison of countries on the basis of the individual-level responses, something that can be done in a very elegant manner using the multilevel LC model.

The Inglehart scale consists of making a first (Y_{ij1}) and second (Y_{ij2}) choice out of a set of four alternatives. Respondents indicate what they think that should be the most and next most important aim of their country for the next ten years. The four alternatives are: 1) "Maintaining order in the nation", 2) "Giving people more say in important government decisions", 3) "Fighting rising prices", and 4) "Protecting freedom of speech". The two observed indicators can be modelled by the LC model for (partial) rankings proposed by Croon (1989), in which $P(Y_{ij1} = s_1, Y_{ij2} = s_2)$ is assumed to be equal to

$$P(Y_{ij1} = s_1, Y_{ij2} = s_2) = \sum_{t=1}^{T} P(X_{ij} = t) \frac{\exp(\beta_{s_1 t})}{\sum_{r=1}^{4} \exp(\beta_{rt})} \frac{\exp(\beta_{s_2 t})}{\sum_{r \neq s_1} \exp(\beta_{rt})}$$

for $s_1 \neq s_2$, and 0 otherwise. This model differs from a standard LC model in that the β parameters are equal for the two items and that the probability of giving the same response on the two items is structurally zero. This is a way to take into account that the preferences remain the same in the both choices and that the second choice cannot be the same as the first one. The β parameters can be interpreted as the class-specific utilities of the alternative concerned. The alternative with the largest positive utility has the highest probability to be selected, and the one with the largest negative value the lowest probability.

The multilevel structure is used to model country differences in the latent distribution. I will again use both a continuous and a discrete mixture distribution in order to be able to see the effect of the specification of this part of the model on the final results. From a substantive point of view, the finite mixture approach yielding a typology of countries is the more natural one in this application.

[Insert Table 5 about here]

Table 5 reports the test results obtained with the EVS data. Again I started estimating models without random effects. Comparison of the BIC values of Models 1, 2, and 3 shows that a two-class model should be preferred for this data set. Note that this is in agreement with the intension of this scale which should make a distinction between materialists and post-materialists. The much lower BIC values obtained with the two-class models with a random log-odds of belonging to the post-materialist class (Models 4-7) indicate that there is considerable between-country variation in the latent distribution. The continuous normal mixing distribution and the three-class finite mixture capture the between-country variation equally well.

[Insert Table 6 about here]

Table 6 reports the parameter estimates obtained with Models 4 and 6. The β_{st} parameters, which are very similar for both models, indicate that class one can be named "materialist" since it has much higher utilities for the "order" and "prices" alternatives and therefore higher probabilities of selecting these alternatives than class two. The other class is named "postmaterialist" because it has higher probabilities of selecting the "more say" and "freedom of speech" alternatives than class one.

There turn out to be large country differences in the probability that a citizen belongs to the postmaterialist class. For example, filling in u_j values of -1.28 and 1.28 in the results of Model 4 yields class-membership probabilities for class two of .09 and 86, respectively. Also the intraclass correlation is very large (.44). These results are confirmed by the nonparametric approach of Model 6 in which we see large differences in the logit of belonging to class two between the three mixture components. The three logits correspond to probabilities of 0.06, 0.53, and 0.97, respectively.

[Insert Table 7 about here]

Table 7 reports the estimated posterior means $\overline{\gamma}_{2j}$ obtained via equations (15) and (16) for each of the 32 countries, as well as the corresponding values for the probability of belonging to latent class two, $P(X_{ij} = 2)$. The countries are ordered from most to least post-materialist based on the results of Model 4. As can be seen, the finite mixture specification yields somewhat more extreme values for $P(X_{ij} = 2)$ and, as can be expected, sharper cut off poins between groups of countries. From a substantive point of view, the results are, however, quite similar. The main difference is that based on the results of Model 4, one might conclude that there are four or five groups of countries.

5 Discussion

Parametric and nonparametric random-coefficient LC models were proposed that make it possible to deal with various types of dependent observations. In addition, an adapted EM algorithm named upward-downward algorithm was presented that makes maximum likelihood estimation feasible.

The new model was illustrated with examples from organizational, educational, and cross-national comparative research. In the first application, I used a random-intercept model to take the nested data structure into account and to estimate the between-group variation in class membership. The second application was similar to a standard random-coefficient multinomial regression model. An important difference was, of course, that we were able to correct for measurement error in the dependent variable. The third application showed that the multilevel LC model can be used as a parsimonious multiple-group model when there are more than a few groups, as well as that it can also be used with slightly more complicated LC models, such as models for ranking data. In each of the empirical examples, there was clear evidence for between-group variation in the latent class distribution.

A problem with the presented parametric approach is that the numerical integration to be performed for parameter estimation can involve summation over a large number of points when the number of random coefficients is increased. It should be noted that the total number of quadrature points equals the product of the number of points used for each dimension. Despite the fact that the number of points per dimension can be somewhat reduced with multiple random coefficients, computational burden becomes enormous with more than 5 or 6 random coefficients. There exist other methods for computing high-dimensional integrals, such as Bayesian simulation and simulated maximum likelihood methods, but these are also computationally intensive. As shown by Vermunt and Van Dijk (2001), these problems can be circumvented by adopting a nonparametric approach in which computation burden is much less dependent on the number of random coefficients. The finite mixture specification is not only interesting on its own, but it can also be used to approximate a multivariate continuous mixture distribution.

References

Agresti Alan, Booth, James G., Hobert, James P., and Caffo, Brian. 2000. "Random-effects modeling of categorical response data." *Sociological Methodology* 30: 27-80.

Bandeen-Roche, Karen J., Miglioretti, Diana L., Zeger, Scott L., and Rathouz, Paul J. 1997. "Latent variable regression for multiple discrete outcomes." *Journal of the American Statistical Association* 92:1375-1386.

Baum, Leonard E., Petrie, Ted, Soules, George and Weiss, Norman 1970. "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains." *Annals of Mathematical Statistics* 41: 164-171.

Bock, R. Darroll 1972. "Estimating item parameters and latent ability when responses are scored in two or more nominal categories." *Psychometrika* 37: 29-51.

Bock, R. Darroll and Aitkin, Murray 1981. "Marginal maximum likelihood estimation of item parameters." *Psychometrika* 46: 443-459.

Bryk, Anthony S., and Raudenbush, Stephen W. 1992. *Hierarchical linear models: application and data analysis*. Newbury Park, CA: Sage Publications.

Clogg, Clifford C., and Goodman, Leo A. 1984. "Latent structure analysis of a set of multidimensional contingency tables." *Journal of the American Statistical Association* 79: 762-771.

Croon, Marcel A. 1989. "Latent class models for the analysis of rankings." Pp 99-121 in *New developments in psychological choice modeling* edited by G. De Soete, H. Feger, and K.C. Klauer. Elsevier Science Publishers.

Dayton, C. Mitchell, and Macready, George B. 1988. "Concomitant-variable latent-class models." *Journal of the American Statistical Association* 83: 173-178.

Doolaard, Simone 1999. "Schools in Change or School in Chain." Ph. D. dissertation, University of Twente, The Netherlands.

Fox, Jean-Paul. and Glas, Cees A.W. 2001. "Bayesian estimation of a multilevel IRT model using Gibbs sampling." *Psychometrika* 66: 269-286.

Goldstein, Harvey 1995. *Multilevel statistical models.* New York: Halsted Press.

Goodman, Leo A. 1974. "The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach." *American Journal of Sociology* 79: 1179-1259.

Hedeker, Donald 1999. "MIXNO: a computer program for mixed-effects nominal logistic regression." *Journal of Statistical Software* 4: 1-92.

Hedeker, Donald, 2002. "A mixed-effects multinomial logistic regression model." *Statistics in Medicine,* in press.

Hedeker, Donald, and Gibbons, Robert D 1996. "MIXOR: A computer program for mixed-effects ordinal regression analysis." *Computer Methods and Programs in Biomedicine* 49: 157-176.

Heinen, Ton 1996. *Latent Class and Discrete Latent Trait Models: Similarities and Differences.* Thousand Oakes: Sage Publications.

Inglehart, Ronald 1977. *The Silent Revolution.* Princeton: Princeton University Press.

Juang, Biing Hwang, and Rabiner, Laurence R. 1991. "Hidden Markov models for speech recognition." *Technometrics* 33: 251-272.

Laird, Nan 1978. "Nonparametric maximum likelihood estimation of a mixture distribution." *Journal of the American Statistical Association* 73: 805-811.

Lazarsfeld, Paul F. 1950. "The logical and mathematical foundation of latent structure analysis & The interpretation and mathematical foundation of latent structure analysis." Pp 362-472 in *Measurement and Prediction* edited by S.A. Stouffer et al. Princeton, NJ: Princeton University Press.

Lenk, Peter J., and DeSarbo, Wayne S. 2000. "Bayesian inference for finite mixture models of generalized linear models with random effects." *Psychometrika* 65: 93-119.

Qu, Yinsheng, Tan, Ming, and Kutner, Michael H. 1996. "Random effects models in latent class analysis for evaluating accuracy of diagnostic tests." *Biometrics*, 52, 797-810.

Rabe-Hesketh, Sophia, Pickles, Andrew, and Skrondal, Anders 2001. "GLLAMM: A general class of multilevel models and a Stata program." *Multilevel Modelling Newsletter* 13: 17-23.

Snijders, Tom A.B. & Bosker, Roel J. 1999. *Multilevel Analysis.* London: Sage Publications.

Stroud, Arthur H. and Secrest. Don. 1966. *Gaussian Quadrature Formulas.* Englewood Cliffs, NJ: Prentice Hall.

Van Mierlo, Heleen, Vermunt, Jeroen K., and Rutte, Cristel, 2002. "Using individual level survey data to measure group constructs: A comparison of items with reference to the individual and to the group in a job design context." Manuscript submitted for publications.

Vermunt, Jeroen K. and Van Dijk, Liesbet 2001. "A nonparametric random-coefficients approach: the latent class regression model." *Multilevel Modelling Newsletter* 13: 6-13.

Vermunt, Jeroen K. and Magidson, Jay 2000. *Latent GOLD User's Manual.* Boston: Statistical Innovations.

Wong, George Y. and Mason, William M. 1985. "Hierarchical logistic models for multilevel analysis." *Journal of the American Statistical Association* 80: 513-524.

Table 1. Testing results obtained with the task-variety data

Model	LC part	Random-coefficients part	LL	# par	BIC
1	1-class	none	-2685	5	5405
2	2-class	none	-2385	11	4844
3	3-class	none	-2375	16	4859
4	2-class	γ_{2j}, 9 nodes	-2368	12	4818
5	2-class	γ_{2m}, 2 classes	-2367	13	4822
6	2-class	γ_{2m}, 3 classes	-2366	15	4835

Table 2. Estimated parameter values of Models 4 and 5 (task-variety data)

	Model 4 (parametric)				Model 5 (nonparametric)			
	t = 1		t = 2		t = 1		t = 2	
Parameter	Value	S.E.	Value	S.E.	Value	S.E.	Value	S.E.
Latent logit mean (γ_2)			0.73	0.15				
Latent logit std. dev. (τ_2)			0.87	0.14				
Latent logit for m=1 (γ_{21})							-0.35	0.20
Latent logit for m=2 (γ_{22})							1.29	0.18
$P(Y_{ij1}=1)$	0.17	0.03	0.54	0.02	0.17	0.03	0.54	0.02
$P(Y_{ij2}=2)$	0.30	0.03	0.71	0.02	0.30	0.03	0.71	0.02
$P(Y_{ij3}=2)$	0.21	0.04	0.97	0.01	0.22	0.04	0.97	0.01
$P(Y_{ij4}=2)$	0.44	0.03	0.83	0.02	0.44	0.03	0.83	0.02
$P(Y_{ij5}=2)$	0.17	0.03	0.94	0.02	0.17	0.03	0.94	0.02

Table 3. Testing results obtained with the mathematical ability data

Model	LC part	Random-coefficients part	LL	# par	BIC
1	1-class	none	-22264	18	44667
2	2-class	none	-20386	37	41055
3	3-class	none	-20175	56	40779
4	4-class	none	-20127	75	40831
5	3-class with predictors	none	-19791	64	40072
6	3-class	γ_{tj}, 9 nodes	-20039	58	40524
7	3-class	γ_{tm}, 2 classes	-20067	59	40586
8	3-class	γ_{tm}, 3 classes	-20039	62	40555
9	3-class	γ_{tm}, 4 classes	-20037	65	40572
10	3-class with predictors	γ_{0tj}, 9 nodes	-19696	66	39899
11	3-class with predictors	γ_{0tm}, 3 classes	-19689	70	39916

Table 4. Estimated parameter values of parametric Model 10 (mathematical ability data)

	t = 2		t = 3	
Parameter	Value	S.E.	Value	S.E.
Intercept/Mean (γ_{0t})	0.82	0.26	-0.70	0.38
SES (γ_{1t})	0.83	0.11	1.36	0.13
ISI (γ_{2t})	1.15	0.12	2.28	0.12
GENDER (γ_{3t})	-0.61	0.18	-1.03	0.21
CITO (γ_{4t})	1.85	0.29	3.08	0.42
Std. dev. (τ_t)	1.11	0.18	2.25	0.27

Table 5. Testing results obtained with the EVS data

Model	LC part	Random-coefficients part	LL	# par	BIC
1	1-class	none	-8392	3	16809
2	2-class	none	-8259	7	16576
3	3-class	none	-8248	11	16587
4	2-class	γ_{2j}, 9 nodes	-8158	8	16382
5	2-class	γ_{2m}, 2 classes	-8191	9	16457
6	2-class	γ_{2m}, 3 classes	-8156	11	16402
7	2-class	γ_{2m}, 4 classes	-8151	13	16409

Table 6. Estimated parameter values of Modes 4 and 6 (EVS data)

	Model 4 (parametric)				Model 6 (nonparametric0			
	t = 1		t = 2		t = 1		t = 2	
Parameter	Value	S.E.	Value	S.E.	Value	S.E.	Value	S.E.
Latent logit mean (γ_2)			-0.27	0.25				
Latent logit std. dev. (τ_2)			1.61	0.20				
Latent logit for m=1 (γ_{21})							-2.75	0.65
Latent logit for m=2 (γ_{22})							0.13	0.20
Latent logit for m=3 (γ_{23})							3.37	1.47
Order (β_{1t})	1.08	0.06	0.32	0.04	1.03	0.07	0.34	0.04
More say (β_{2t})	0.07	0.05	0.27	0.04	0.07	0.05	0.26	0.04
Prices (β_{3t})	0.41	0.07	-0.66	0.06	0.40	0.07	-0.67	0.07
Freedom of speech (β_{4t})	-1.56	0.14	0.08	0.05	-1.50	0.15	0.06	0.06

Table 7. Posterior means of random effects and the corresponding proportion of postmaterialists for the 32 countries (Models 4 and 6, EVS data)

	Model 4 (parametric)		Model 6 (nonparametric)	
Country	γ_{tj}	$P(X_{ij}=2)$	γ_{tj}	$P(X_{ij}=2)$
Italy	3.14	0.96	3.37	0.97
Sweden	2.99	0.95	3.37	0.97
Denmark	2.96	0.95	3.37	0.97
Austria	2.83	0.94	3.37	0.97
Netherlands	2.81	0.94	3.37	0.97
Croatia	1.67	0.84	3.29	0.96
Belgium	1.43	0.81	3.36	0.97
Greece	1.35	0.79	0.41	0.60
France	1.35	0.79	0.13	0.53
Spain	1.24	0.78	0.14	0.54
Northern Ireland	1.15	0.76	0.14	0.54
Ireland	1.06	0.74	0.16	0.54
Luxembourg	0.02	0.51	0.13	0.53
Slovenia	-0.16	0.46	0.13	0.53
Czechia	-0.23	0.44	0.13	0.53
Iceland	-0.27	0.43	0.13	0.53
Finland	-0.27	0.43	0.13	0.53
West Germany	-0.27	0.43	0.13	0.53
Portugal	-0.28	0.43	0.13	0.53
Romania	-0.29	0.43	0.13	0.53
Malta	-0.29	0.43	0.11	0.53
East Germany	-0.31	0.42	0.12	0.53
Bulgaria	-0.31	0.42	0.12	0.53
Lithuania	-1.83	0.14	-2.71	0.06
Latvia	-1.87	0.13	-2.74	0.06
Poland	-1.91	0.13	-2.75	0.06
Estonia	-1.96	0.12	-2.75	0.06
Belarus	-2.21	0.10	-2.75	0.06
Slovakia	-2.24	0.10	-2.75	0.06
Hungary	-2.39	0.08	-2.75	0.06
Ukraine	-2.78	0.06	-2.75	0.06
Russia	-3.75	0.02	-2.75	0.06

Figure 1: Graphical model associated with a LC model with a random-effect in the latent class distribution.