Mixture models for multilevel data sets

Jeroen K. Vermunt

Department of Methodology and Statistics, Tilburg University

The Netherlands

1. Introduction

Latent class (LC) and mixture models are nowadays part of the standard statistical toolbox of researchers in applied fields such as sociology, psychology, marketing, biology, and medicine. In most of their applications, the aim is to cluster units into a small number of latent classes or mixture components (McLachlan and Peel, 2000). This clustering can be based on a set of categorical response variables as in the traditional LC model (Goodman, 1974), on a set of continuous items as in a latent profile model (Lazarsfeld and Henry, 1968; Vermunt and Magidson, 2002) and mixture factor analysis (McLachlan and Peel, 2000; Yung, 1997), on a set of repeated measures as is mixture growth models (Nagin, 1999; Muthén, 2004; Vermunt, 2007b) and mixture (latent) Markov models (Van der Pol and Langeheine, 1990; Vermunt, Tran, and Magidson, 2008), or on other types of two-level data sets, such as from experiments in which individuals are confronted with multiple experimental conditions (Wedel and DeSarbo, 1994) and from studies in which multiple individuals are nested within higher-level units (Aitkin, 1999; Vermunt and Van Dijk, 2001; Kalmijn and Vermunt, 2007). Whereas the main application of LC models is clustering, restricted LC models and LC models with multiple latent variables can be also used for scaling type of applications similar to IRT and factor analysis (Heinen, 1996; Vermunt, 2001; Magidson and Vermunt, 2001).

Recently, a multilevel extension of the LC model was proposed (Vermunt, 2003, 2008a). It can, for example, be used when individuals with multiple item responses or repeated measurements are nested within groups (Vermunt, 2003, 2005, 2008a; Bijmolt, Paas, Vermunt, 2004), when multivariate repeated responses are nested within individuals (Vermunt, Tran, and Magidson, 2008), as well as with three-mode data sets on individuals observed with different measures in different situations (Vermunt, 2007; Bouwmeester, Vermunt, and Sijtsma, 2007) and three-level data sets (Vermunt, 2004, 2008a). As in standard LC models, for the lower level units, the main goal will usually be to build a meaningful cluster model. One variant of this hierarchical LC model yields also a clustering of higher-level units by assuming that these belong to higher-level latent classes which differ in either lower-level responses or lower-level class membership probabilities. Another variant makes use of random effects to capture higher-level variation in LC model parameters, especially in lower-level class membership probabilities.

The aim of this chapter is threefold. The first is to explain the relationship between LC analysis and multilevel regression analysis techniques. It will be shown that LC models can be conceptualized as models for two-level data sets in which parameters vary randomly across level-2 units. Whereas in multilevel regression analysis this variation is modeled by assuming that parameters come from a particular continuous distribution (typically normal), which is equivalent to introducing one or more continuous latent variables in the model, in LC analysis variation is modeled using discrete latent variables (Aitkin, 1999; Vermunt and Van Dijk, 2001).

Our second aim is to introduce the multilevel extension of the LC model proposed by Vermunt (2003, 2008a), which is a model for univariate three-level data sets and multivariate two-level data sets. The multilevel LC model uses either continuous or discrete latent variables at the higher level.

Third, we discuss other kinds of mixture models for these types of multilevel data sets by connecting multilevel LC analysis to the general latent variable modeling framework described by Skrondal and Rabe-Hesketh (2004). This framework integrates factor analytic and random effects models, as well as models with continuous and discrete latent variables (see also Asparouhov and Muthén, 2008, and Vermunt, 2008b). We present a nine-fold classification of latent variable models, where eight types can be labeled multilevel mixture models. Most of these models are implemented in software packages such as GLLAMM (Rabe-Hesketh, Skrondal, and Pickles, 2004), Mplus (Muthén and Muthén, 2008), and Latent GOLD (Vermunt and Magidson, 2008).

2. Two-level data sets: the standard latent class model

Whereas the traditional LC model was developed for the analysis multivariate response data sets (Goodman, 1974; Lazarsfeld and Henry, 1968), LC analysis can also been used for the analysis of two-level data sets (Aitkin, 1999; Vermunt and Van Dijk, 2001; Wedel and DeSarbo, 1994), such as from students nested within schools and from longitudinal studies in which repeated measurement are nested within persons. However, when the responses are of the same type – for example, all binary or all continuous – we can also conceptualize the traditional LC model as a model two-level data; that is, by treating the single-level multivariate responses (say on questionnaire or test items) as two-level univariate responses, as item responses nested within individuals. Note that this is similar to IRT modeling using multilevel techniques by treating multiple item responses as multiple observations nested within individuals (see De Boeck and Wilson, 2004). A similar connection has also been established between factor analysis and multilevel linear regression analysis (Hox, 2002).

Using the typical multilevel analysis notation, let y_{ij} denote the response of level-1 unit *i* belonging to level-2 unit *j*, n_j the total the number of level-1 units within level-2 unit *j*, and Y_j the complete response vector of unit *j*. This could thus also be the n_j item responses of person *j*, where *i* refers to a particular item and where n_j takes on the same for all persons. We refer to a particular latent class by the symbol *c* and to the number of latent classes by *C*. To stress the similarity between the discrete latent variable in a LC model and random effects in a multilevel regression model, we use the symbol *u* to refer to the latent class membership of unit *j*. More specifically, u_{jc} is one of *C* indicator variables, which take on the value 1 if level-2 unit *j* belongs to latent class *c* and 0 otherwise. Because classes are exhaustive and mutually exclusive, exactly one of the *C* class indicators u_{jc} equals 1 and the others equal 0. The vector of class indicators is denoted by U_j. Using this notation, a standard LC model can be defined as follows:

$$\begin{aligned} f(Y_j) &= \sum_{c=1}^{C} P(u_{jc}=1) f(Y_j \mid u_{jc}=1) \\ &= \sum_{c=1}^{C} P(u_{jc}=1) \prod_{i=1}^{n_j} f(y_{ij} \mid u_{jc}=1) \end{aligned} \quad (1)$$

As can be seen, the LC model is a model for Y_j, the full response vector of level-2 unit *j*. The model equation shows the two basic assumptions of a LC model. The first one is that the density of Y_j, $f(Y_j)$, is a weighted sum of class specific densities $f(Y_j \mid u_{jc} = 1)$, where the class proportion $P(u_{jc} = 1)$ serve as weights. In other words, a level-2 unit has a certain prior probability of belonging to class *c*, and conditional on belonging to class *c* it has a certain probability of giving responses Y_j. The second basic assumption is that the level-1 responses are independent of one another given a level-2 unit's class membership. This assumption is typically referred as the local independence assumption. Note that this assumption is also made in factor analytic and random effects models, in which responses are assumed to be independent conditional on a unit's latent factor scores and random effects values, respectively.

The specific form chosen for the conditional density $f(y_{ij} \mid u_{jc} = 1)$ depends on the scale type of the response variable. Examples are Bernoulli for binary responses, normal for continuous responses, and Poisson for counts. What we are typically interested in are the expected values of these conditional distributions, denoted by $E(y_{ij} \mid u_{jc} = 1)$, which can be binomial proportions, normal means, Poisson rates, etc. These class-specific expected values can be parameterized using a generalized linear model; i.e., using a linear model after applying the appropriate link function $g[\cdot]$. Let us assume that we are dealing with a traditional LC model, which means that the level-1 observations are n_j questionnaire items. A regression model for the $n_j \cdot C$ conditional means $E(y_{ij} \mid u_{jc} = 1)$ can be formulated as follows:

$$g[E(y_{ij} \mid u_{jc} = 1)] = \beta_0 + \beta_i + \lambda_{0c} + \lambda_{ic}.$$

This (unrestricted) model contains an intercept (β_0), item effects (β_i), main effects for the classes (λ_{0c}), and item-class interactions (λ_{ic}). Note that this model contains $1 + n_j + C + n_j \cdot C$ unknown parameters, which means that $1 + n_j + C$ identification restrictions should be imposed on the regression coefficients, for example, by using dummy coding where the parameters for the first item and the first class are fixed to zero, or by fixing the β_0, β_i, and λ_{0c} parameters to zero.

To show how the same regression model for the n_j item responses can be formulated in the more typical multilevel analysis notation, let X_j and Z_j be two (identical) design matrices with n_j rows and $n_j + 1$ columns:

$$X_j = Z_j = \begin{bmatrix} 1 & 1 & 0 & . & 0 \\ 1 & 0 & 1 & . & 0 \\ 1 & 0 & 0 & . & 0 \\ . & . & . & . & . \\ 1 & 0 & 0 & . & 1 \end{bmatrix} .$$

The first column of X_j and Z_j contains the 1s for the intercept term and the remaining columns contain dummies for the item effects. Using these design matrices and conditioning on the vector U_j rather than on $u_{jc} = 1$, the regression model for y_{ij} can also be written as follows:

$$g[E(y_{ij} \mid U_j, X_{ij,} Z_{ij})] = \sum_{p=0}^{P} \beta_p x_{pij} + \sum_{q=0}^{Q} \left(\sum_{c=1}^{C} \lambda_{qc} u_{jc} \right) z_{qij} . \quad (2)$$

Though is equation is already very similar to a two-level regression model, the similarity becomes even clearer if we define $u_{qj}^{*} = \sum_{c=1}^{C} \lambda_{qc} u_{jc}$; that is,

$$g[E(y_{ij} \mid U_j, X_{ij,} Z_{ij})] = \sum_{p=0}^{P} \beta_p x_{pij} + \sum_{q=0}^{Q} u_{qj}^{*} z_{qij} .$$

This equation shows that a LC model can be seen as a model with random effects u_{qj}^{*}. A difference compared to a standard random-effects model is that rather than assuming that the random effects come from a *Q*-dimensional multivariate normal distribution, here we assume that these take on only *C* different values, where each value λ_{qc} occurs with probability $P(u_{jc}=1)$. Both the values of the random effects (the locations) and the associated probabilities are free parameters to be estimated. The mean of u_{qj}^{*} and covariance between u_{qj}^{*} and $u_{q'j}^{*}$ can be obtained as $\overline{u}_q^{*}=\sum_{c=1}^{C}\lambda_{qc}P(u_{jc}=1)$ and $\sigma_{u_q u_{q'}}=\sum_{c=1}^{C}\lambda_{qc}\lambda_{q'c}P(u_{jc}=1)-\overline{u}_q^{*}\,\overline{u}_{q'}^{*}$ (Aitkin, 1999; Vermunt and Van Dijk, 2001).

In the above example, we used design matrices X_j and Z_j with a very specific structure, with an intercept and a set of item effects. However, as in two-level regression models, these two matrices may also contain the values of other types of predictors. More specifically X_j may contain level-1 predictors, level-2 predictors, and cross-level interactions, and Z_j may contain level-2 predictors and cross-level interactions. This yields what is often referred to as a LC or mixture regression model (Aitkin, 1999; Vermunt and Van Dijk, 2001; Wedel and Kamakura, 1994), which is a regression model for two-level data sets.

In LC models, level-2 predictors cannot only be used in the regression model for the response variable y_{ij}, but can also be used to predict the class membership probabilities. Typically a multinomial logit model is used for this purpose:

$$\text{logit}[P(u_{jc}=1\mid X_j^{(2)})]=\sum_{r=0}^{R}\gamma_{rc}x_{rj}^{(2)}\,, \tag{3}$$

where the first column of $X_j^{(2)}$ ($r = 0$) defines the constant term. For identification, one constraint has to be imposed on γ_{rc} for each r, for example, $\gamma_{r1} = 0$.

It is important to note that also in standard two-level regression models one may define regression models for the latent variables representing the random effects, as is usually done in the hierarchical model specification of the multilevel model. However, it is always possible to substitute the random effects in the model for the response variable by their regression equations, yielding the well-known mixed model formulation of the multilevel model. Such a substitution is not possible in LC regression analysis.

Whereas the standard LC model is a latent variable model with a nominal latent variable, it is also possible to defined LC models with a latent variable with ordered categories. Such an ordinal specification is obtained by using a single u_j with numeric scores instead of *C* class indicators u_{jc} ((Heinen, 1996; Vermunt, 2001), which implies replacing the term $\sum_{c=1}^{C} \lambda_{qc} u_{jc}$ in equation (2) by $\lambda_q u_j$. In a three-class model, the class locations u_j could, for example be -1, 0, and 1 or 0, 0.5, and 1.

The above model formulations can easily be adapted for LC models with more than one nominal or ordinal latent variable. For example, the LC model with multiple ordinal latent variables proposed by Magidson an Vermunt (2001) can be defined by including a term $\sum_{\ell=1}^{L} \lambda_{q\ell} u_{\ell j}$ in equation (2), where $u_{\ell j}$ is one of *L* ordinal latent variables. Similarly, a model with *L* nominal latent variables can be defined by setting up a series of dummies for each latent variable and using the term $\sum_{\ell=1}^{L} \sum_{c=1}^{C_\ell} \lambda_{q\ell c} u_{\ell jc}$ in equation (2). In models with multiple latent variables, one will typically restrict the joint probability density of the L latent variables, for example, using a log-linear (path) model (Hagenaars, 1993; Vermunt 1997) or a latent Markov structure (Van de Pol and Langeheine, 1990).

3. Three-level data sets: the multilevel latent class model

Let us now expand the LC model for the situation in which we have either a three-level univariate response or a two-level multivariate response data set. Note that by conceptualizing multivariate responses as nested univariate responses, the latter can also be seen as three-level data sets, which is what we will do here. The extension of the LC model discussed here yields what Vermunt (2003, 2008a) called multilevel LC analysis.

To accommodate the additional hierarchical level, two modification of the notation introduced in the previous sections are needed: an index k is used to refer to a particular level-3 unit and, whenever necessary, a superscript (1), (2), or (3) is used to denote whether we are referring to a level-1, level-2, or level-3 quantity. For example, level-1 responses are now denoted by y_{ijk}, a level-2 response vector by Y_{jk}, level-2 class indicators by $u_{jkc}^{(2)}$, and a level-2 vector of class indicators by $U_{jk}^{(2)}$.

The main difference compared to the two-level case discussed in the previous section is that a multilevel LC model contains either continuous random effects or a discrete latent variable (=discrete random effects) at level three. These random effects pick up variation in LC model parameters across level-3 units. Below, we first discuss the situation in which level-3 heterogeneity is modeled using discrete random effects, as well as two important special cases of this specification. Then we discuss the multilevel LC model with continuous random effects. The third part of this section introduces other types of multilevel mixture models; that is, models with discrete random effects at level 3 but which are not necessarily LC models at level 2.

3.1 Models with discrete random effects at level three

Let us first look at the situation in which the heterogeneity at the highest level is modeled by assuming that level-3 units belong to one of D latent classes. The level-3 class membership is denoted using indicator variables $u_{kd}^{(3)}$, which take on the value 1 if unit k belongs to class d and 0 otherwise. The vector of level-3 class indicators is denoted by $U_k^{(3)}$. The corresponding multilevel extension of the LC model is as follows:

$$f(Y_{jk} \mid u_{kd}^{(3)}=1)=\sum_{c=1}^{C} P(u_{jkc}^{(2)}=1 \mid u_{kd}^{(3)}=1)\prod_{i=1}^{n_{jk}^{(2)}} f(y_{ijk} \mid u_{jkc}^{(2)}=1, u_{kd}^{(3)}=1)\,. \qquad (4)$$

As can be seen, the "only" modification compared to the standard LC model described in equation (1) is that both the level-2 class proportions $P(u_{jkc}^{(2)}=1 \mid u_{kd}^{(3)}=1)$ and the class-specific densities $f(y_{ijk} \mid u_{jkc}^{(2)}=1, u_{kd}^{(3)}=1)$ may depend on $U_k^{(3)}$. It is important to note that a multilevel LC model is actually a model for Y_k, the full response vector of higher-level unit k; that is,

$$f(Y_k)=\sum_{d=1}^{D} P(u_{kd}^{(3)}=1)\prod_{j=1}^{n_k^{(3)}} f(Y_{jk} \mid u_{kd}^{(3)}=1) \qquad (5)$$

where $f(Y_{jk} \mid u_{kd}^{(3)}=1)$ was defined in equation (4). It can easily be seen that equation (5) defines a LC model for the higher level units, which is very similar to equation (1), the equation for a standard latent model. Equation (5) shows that groups are assumed to belong to one of D latent classes, as well as that level-2 observations are assumed to be independent of one another conditional on the level-3 class membership.

The fact that $P(u_{jkc}^{(2)}=1 \mid u_{kd}^{(3)}=1)$ and $f(y_{ijk} \mid u_{jkc}^{(2)}=1, u_{kd}^{(3)}=1)$ depend on $u_{kd}^{(3)}$ can also be expressed via the regression models for $u_{jkc}^{(2)}$ and y_{ijk}. These will differ from the ones in equations (2) and (3) in that they may now contain terms for $u_{kd}^{(3)}$. In addition, a logistic regression model may be specified for $u_{kd}^{(3)}$ itself. We use (sometimes double) superscripts to distinguish the different parameters sets and design matrices, where the first index refers to the level of the dependent variable in the equation concerned and the second, if present, to the level of the random effect. The three regression equations defining the multilevel LC model are:

$$g[E(y_{ijk} \mid U_{jk}^{(2)}, U_k^{(3)}, X_{ijk}^{(1)}, Z_{ijk}^{(1)})] = \sum_{p=0}^{P} \beta_p x_{pijk}^{(1)} + \sum_{q=0}^{Q^{(2)}} \left(\sum_{c=1}^{C} \lambda_{qc}^{(1,2)} u_{jkc}^{(2)} \right) z_{qijk}^{(1,2)} + \sum_{q=0}^{Q^{(3)}} \left(\sum_{d=1}^{D} \lambda_{qd}^{(1,3)} u_{kd}^{(3)} \right) z_{qijk}^{(1,3)} \quad (6)$$

$$\text{logit}[P(u_{jkc}^{(2)}=1 \mid U_k^{(3)}, X_{jk}^{(2)}, Z_{jk}^{(2)})] = \sum_{r=0}^{R} \gamma_{rc}^{(2)} x_{rjk}^{(2)} + \sum_{s=0}^{S} \left(\sum_{d=1}^{D} \lambda_{scd}^{(2,3)} u_{kd}^{(3)} \right) z_{sjk}^{(2,3)} \quad (7)$$

$$\text{logit}[P(u_{kd}^{(3)}=1 \mid X_k^{(3)})] = \sum_{t=0}^{T} \gamma_{td}^{(3)} x_{tk}^{(3)} . \quad (8)$$

The terms $\sum_{c=1}^{C} \lambda_{qc}^{(1,2)} u_{jkc}^{(2)}$, $\sum_{d=1}^{D} \lambda_{qd}^{(1,3)} u_{kd}^{(3)}$, and $\sum_{d=1}^{D} \lambda_{scd}^{(2,3)} u_{kd}^{(3)}$ have a similar (discrete) random effects interpretation as was explained in the previous section. As can be seen, as in the 2-level model, the level-2 classes may capture variation in the parameters of the model for the response variable. The level-3 classes may capture variation in the parameters of the response model, as well as in the parameters of the regression model for the level-2 classes.

When the model does not contain predictors and when, as in the example used in the previous section, the design matrices are setup to yield intercepts and item parameters, the three regression equations can be written in a simpler form; that is,

$$g[E(y_{ijk} \mid u_{jkc}^{(2)} = 1, u_{kd}^{(3)} = 1)] = \beta_0 + \beta_i + \lambda_{0c}^{(1,2)} + \lambda_{ic}^{(1,2)} + \lambda_{0d}^{(1,3)} + \lambda_{id}^{(1,3)},$$

$$\text{logit}[P(u_{jkc}^{(2)} = 1 \mid u_{kd}^{(3)} = 1)] = \gamma_{0c}^{(2)} + \lambda_{0cd}^{(2,3)},$$

$$\text{logit}[P(u_{kd}^{(3)} = 1)] = \gamma_{0d}^{(3)}.$$

which yields the more typical LC analysis notation.

Let us look at two more restricted special cases of the model defined in equations (4) to (8). The first special case is obtained by assuming that groups differ in the lower-level class membership probabilities but have the same response variable densities. This implies that $f(y_{ijk} \mid u_{jkc}^{(2)} = 1, u_{kd}^{(3)} = 1) = f(y_{ijk} \mid u_{jkc}^{(2)} = 1)$ or, equivalently, that the term containing the $\lambda_{qd}^{(1,3)}$ parameters is excluded from equation (6). This is the special case used by Vermunt (2003, 2007). The basic idea is that the model part linking the lower-level class memberships to the responses is the same for all groups, which is conceptually similar to saying that there is measurement equivalence across higher-level units. Groups may, however, differ with respect to the lower-level class membership probabilities of their members, as well as with respect to covariate effects on these class membership probabilities. These differences across group-level classes are defined with the second term in equation (7).

The second special case is the opposite from the first. It assumes that response densities depend on the group-level class membership, but lower-level class membership not. This implies that $P(u_{jkc}^{(2)} = 1 \mid u_{kd}^{(3)} = 1) = P(u_{jkc}^{(2)} = 1)$ or that the term containing the $\lambda_{scd}^{(2,3)}$ parameters is omitted from equation (7). This model is very similar to a standard regression model in which the variation in the responses is decomposed into independent parts (Goldstein, 2003). The model is also similar to the multilevel factor analysis model proposed by Muthén (1994) and described by Hox (2002), in which the variation in a multivariate response vector is attributed to common latent factors at two levels of a hierarchical structure.

It should be noted that in three-level regression modeling with continuous random effects these two special cases yield equivalent models: regressing a lower-level random effect on a higher-level random effect is the same as a using the terms concerned in the model for the response variables (Goldstein, 2003; Hox, 2002). In the case of a multilevel factor analysis, the first specification is a restricted special case of the second one, which is obtained by equating the factor loadings across levels. In other words, indicating that the lower-level factor means vary randomly across higher-level units is equivalent to having a set of higher-level factors with the same loadings as the lower-level factors. Despite of the conceptual similarities of these models with the multilevel LC model, these equivalences do not apply to the latter.

The full model is conceptually similar to a three-level model in which level-2 and level-3 random effects are correlated, which is a specification that is seldom used in multilevel regression analysis. Another specific aspect of the multilevel LC model is that also level-1 variances (which are free parameters in linear models with normally distributed residuals) can be allowed to differ across lower- and/or higher-level classes. In other words, level-2 and level-3 units may randomly differ with respect to their level-1 variances. There exists no equivalent specification for this in standard three-level regression analysis. A last specific feature I would like to mention is that interactions between level-2 and level-3 class effects can easily be included in the model; that is, by adding terms containing the product $u_{jkc}^{(2)} \cdot u_{kd}^{(3)}$ to equation (6). Also for such interactions there is no equivalence in standard three-level regression analysis. Though this seems to be a somewhat exotic option, this is clearly not the case. A possible application is the investigation of item bias, where not only item intercepts but also lower-level class effects on items may differ across higher-level classes. The latter would be similar to allowing that factor loadings differ randomly across groups, which is not possible in (standard) multilevel factor analysis.

3.2 Models with continuous random effects at level three

Rather than using discrete random effects, it is also possible to use a more standard specification with continuous random effects. Vermunt (2003, 2005) proposed a multilevel LC model in which the class-membership probabilities of lower-level units vary randomly across level-3 units. Using as much as possible the same notation as above, but now with $u_{sk}^{(3)}$ representing the *s*th random effect and $U_k^{(3)}$ the vector of random effects, we can write the model as follows:

$$f(Y_{jk} \mid U_k^{(3)}) = \sum_{c=1}^{C} P(u_{jkc}^{(2)} = 1 \mid U_k^{(3)}) \prod_{i=1}^{n_{jk}^{(2)}} f(y_{ijk} \mid u_{jkc}^{(2)} = 1) . \tag{9}$$

$$f(Y_k) = \int_u f(u_k^{(3)}) \prod_{j=1}^{n_k^{(3)}} f(Y_{jk} \mid u_{kd}^{(3)} = 1) \, du_k^{(3)} \tag{10}$$

with regression equations

$$g[E(y_{ijk} \mid U_{jk}^{(2)}, X_{ijk}^{(1)}, Z_{ijk}^{(1)})] = \sum\nolimits_{p=0}^{P} \beta_p x_{pijk}^{(1)} + \sum\nolimits_{q=0}^{Q^{(2)}} \left(\sum\nolimits_{c=1}^{C} \lambda_{qc}^{(1)} u_{jkc}^{(2)} \right) z_{qijk}^{(1)} \tag{11}$$

$$\text{logit}[P(u_{jkc}^{(2)} = 1 \mid U_k^{(3)}, X_{jk}^{(2)}, Z_{jk}^{(2)})] = \sum\nolimits_{r=0}^{R} \gamma_{rc}^{(2)} x_{rjk}^{(2)} + \sum\nolimits_{s=0}^{S} \lambda_{sc}^{(2)} u_{sk}^{(3)} z_{sjk}^{(2)} \tag{12}$$

Note that this specification assumes measurement equivalence; that is, the parameters in the LC model for the response variable(s) do not vary across groups. Groups differ with respect of their level-2 class membership probabilities, which is specified using a random effects multinomial logistic regression model of the type proposed by Hedeker (2003) for the discrete unobserved variable $U_{jk}^{(2)}$. This model uses a lower-dimensional representation of the *S*+1 random effects for each of the *C* latent classes; that is, it uses the "factor analytic" constraint $u_{sck}^{*(3)} = \lambda_{sc}^{(2)} u_{sk}^{(3)}$, which actually implies that the random effects for a particular *s* are perfectly correlated across latent classes (across values of *c*). Without covariates, equation (12) reduces to a variance decomposition of the class membership logits. A more detailed description of this type of multilevel LC model is provided by Vermunt (2005).

3.3 Other types of multilevel mixture models

The term multilevel LC or mixture model was used so far for the situation in which we have a LC model for level-2 observations combined with an additional hierarchical level, where this third level is dealt with using either continuous or discrete random effects. However, looking at equation (5) which defines a mixture model for level-3 units, one can think of other types of multilevel mixture models; that is, models with latent classes at level three and continuous latent variables (or random effects) at level two. An example is a variant of the mixture factor analysis model where mixture components are not formed by individuals but by groups (Varriale, 2008). Moreover, there is nothing that prevents applying the same logic of hierarchically structured mixture models to situations with more than three hierarchical levels.

{INSERT TABLE 1 ABOUT HERE}

Using such a more general perspective, we get into a general latent variable modeling framework described by Skrondal and Rabe-Hesketh (2004) and expanded in certain ways by Vermunt (2008b) and Muthén (this volume). Table 1 shows the nine-fold classification of latent variable models for three-level data sets based on the scale types of the latent variables at level 2 and level 3. At each level one may have continuous latent variables (or random effects), discrete latent variables, or a combination of these. All models except type A1 could be called multilevel mixture models; that is, models with latent classes at either one or at two levels.

Multilevel factor and IRT models (Fox and Glas, 2001; Goldstein and Brown, 2002; Grilli and Rampichini, 2007; Muthén, 1994), as well as three-level random effects regression models belong to category A1. The multilevel latent models described above belong, depending on whether the level-3 random effects are treated as continuous or discrete, to either the B1 or B2 type. By introduction a continuous latent variable at level two, which is a method for dealing with local dependencies between items in LC analysis, one would obtain a model of type C1 or C2. An example of a model of type C1 is the multilevel variant of the factor mixture model (Allua, 2007). Vermunt (2008b and Varriale (2008) used type A2 models to define multilevel mixture IRT and factor analytic models, respectively. Palardy and Vermunt (2008) used type A2 and A3 models in the context of multilevel mixture growth modeling.

4. Estimation, model selection, and sample size issues

Vermunt (2003, 2004, 2005) demonstrated how to obtain maximum likelihood (ML) estimates of the parameters of multilevel LC models by means the EM algorithm. For this purpose, a (necessary) modification of the E step of the EM algorithm was developed which was called an upward-downward algorithm. This procedure, as well as a Newton-Raphson algorithm with numerical second derivatives based on analytical first derivatives, is implemented in the Latent GOLD software package (Vermunt and Magidson, 2008). The appendix provides various examples of Latent GOLD syntax files. Also Mplus (Muthén and Muthén, 2008) can be used to estimate (most of) the models described in this chapter. The Mplus manual is not very explicit about the estimation methods and algorithms which are used, but it seems to use similar procedures as Latent GOLD. The GLLAMM program (Rabe-Hesketh, Skrondal, and Pickles, 2004) can deal with the situation in which the latent variables at both levels are discrete (or both continuous) and in which only the responses depend on the higher-level class membership (special case number two described in section 3.1). Optimization of the log-likelihood is performed using the Stata ML routines. Each of these three packages has options for obtaining robust standard errors as well as for dealing with missing values and complex sampling designs.

Model selection is already a rather complicated issue in standard LC analysis, but becomes even more complex in multilevel LC models, especially when the level-3 heterogeneity is modeled using level-3 latent classes. We then not only need to decide about the required number of latent classes at level 2 (the value of C), but also about the number of classes at level 3 (the value of D). In principle, in multilevel LC analysis the same types of model selection measures can be used as the ones that are used in standard LC analysis, with information criteria such as AIC, BIC, and AIC3 as the most popular ones. However, the use of BIC is somewhat problematic because it contains the sample size in its formula, and is not fully clear what sample size to use in the BIC formula in a multilevel analysis. Note that this is a problem for multilevel analysis in general, and thus not specific for multilevel LC analysis. A recent simulation study by Lukociene and Vermunt (2008) has shed some light on this issue: when deciding about the number of classes at the higher-level it is better to use the higher-level sample size in the BIC formula instead of the lower-level sample size.

Multilevel LC models have been applied with very different level-1, level-2, and level-3 sample sizes. Although little research has been done on this topic so far, some guidelines can be provided on sample size requirement. What should be understood is that in these types of models the sample size at a particular level may affect not only sampling fluctuation but also the separation of the classes at higher levels, which is similar to the reliability of the measurement of a continuous latent variable. The total level-2 sample size and the level-3 sample size affect the sampling fluctuation in the level-2 and level-3 parameters, respectively. The required level-2 and level-3 sample size depends strongly on the separation of the level-2 and level-3 classes, respectively. But this separation is itself affected by the level-1 sample size for the level-2 classes (the longer a test the more certain we can be about a person's class membership), and by the level-2 sample size for the level-3 classes (the more level-2 units, the more certain we can be about a level-3 unit's class membership). This shows that sample size requirements for one level may depend on the sample size at another level; for example, because of a better separation between level-3 classes, the required level-3 sample size is smaller with larger level-2 and level-1 sample sizes. It should be noted that there are other factors affecting the separation between classes, and thus the required sample size. One of these is how different latent classes are (the size of the λ parameters appearing in the above formulae): smaller samples are need with more dissimilar classes. Another factor is the scale type of the response variable: with continuous normal responses or Poisson counts smaller samples are needed than with the same number of dichotomous responses because the former are more informative about differences between classes.

5. Applications

This section presents two applications of the multilevel LC models described in the previous section. The first is a typical three-level regression application, and concerns a data set containing repeated measurements from a longitudinal survey with individuals nested within regions. The second example uses a data set as is typically analyzed using cluster analysis or factor analysis, with the complicating factor that individuals (children) are nested within groups (families); that is, this is an example of an analysis of a two-level data set with multiple continuous items.

For our analysis we used version 4.5 of the Latent GOLD program (Vermunt and Magidson, 2008). The Appendix presents examples of Latent GOLD syntax files. For model selection, we used two versions of BIC: BIC(2) based on the level-2 sample size and BIC(3) based on the level-3 sample size. When the two fit measures disagree with respect to the required number of level-2 classes, we select the model with the lowest BIC(2). Similarly, when disagreement concerns the number of level-3 classes, we select the model that is preferred by BIC(3).

5.1 Three-level mixture regression analysis

The first application uses a data set from the data library of the Centre of Multilevel Modelling, University of Bristol (http://www.cmm.bristol.ac.uk/learningtraining/multilevel-m-support/datasets.shtml). The data consist of 264 participants in the 1983 to 1986 yearly waves from the British Social Attitudes Survey (McGrath and Waterton, 1986). It is a three-level data set: Individuals are nested within districts and time points are nested within individuals. The total number of level-3 units (districts) is 54.

The dependent variable is the number of yes responses on seven yes/no questions as to whether it is a woman's right to have an abortion under a specific circumstance. Because this variable is a count with a fixed total, it is most natural to work with a binomial error function and a logit link. Individual-level predictors in the data set are religion, political preference, gender, age, and self-assessed social class. In accordance with the results of Goldstein (2003), we found no significant effects of gender, age, self-assessed social class, and political preference. Therefore, we did not use these predictors in the further analysis. The predictors that were used are the level-1 predictor year of measurement (1983, 1984, 1985, and 1986) and the level-2 predictor religion (Roman Catholic, Protestant, Other, and No religion). Because there was no evidence for a linear time effect, we used time as a categorical predictor in the regression model.

Vermunt (2004) used this data set to illustrate the similarity between three-level mixture regression models with intercept variation across latent classes and standard three-level random-intercept models. Here, I present a more extended analysis in which among others slopes are allowed to vary across level-2 and level-3 classes.

Our baseline model is a three-level mixture regression model of the form described in equations (4)-(8). More specifically, we specified models with a fixed intercept and 6 fixed slopes (three for time and three for religion), a random intercept and random slopes for the time effects at level 2, and a random intercept and random slopes for the religion effects religion at level 3. This implies that $X_{ijk}^{(1)}$, $Z_{ijk}^{(1,2)}$, and $Z_{ijk}^{(1,3)}$ contain 7, 4, and 4 columns, respectively. Furthermore, we assume that the level-2 class membership does not depend on the level-3 class membership (this is the second special case discussed in section 3.1) nor on covariates. This means that equation (7) contains only an intercept. Also equation (8) contains only an intercept since we have no level-3 predictors.

{INSERT TABLE 2 ABOUT HERE}

Table 2 reports the log-likelihood (LL) value, the number of parameters (Npar), the BIC(2) value, and the BIC(3) value for models with 1 to 5 level-2 classes and 1 to 3 level-3 classes. The fact that models with C>1 (for constant D) have lower BIC values than models with C=1 shows that there is evidence for level-2 variation in intercept and/or time slopes. A similar conclusion can be drawn for the level-3 variation in intercept and/or religion slopes since models with D>1 perform better according to the BIC statistics than the ones with D=1. BIC(2) and BIC(3) select the same model as the best one, namely the model with C=4 and D=2.

As a next step, we defined five alternative models to test specific aspects of our baseline model with C=4 and D=2. Two more restricted models omit the random time and religion slopes, respectively. Three more extended models are estimated which add a level-3 random time effects, a direct effect of level-3 class on level-2 class, and an interaction effect between level-2 and level-3 class in the model for the response variable. The fits measures in Table 2 indicate that the two more restricted models perform worse than the baseline model, which means that the level-2 and level-3 variation in the time and religion slopes are significant. The models with level-3 variation in the time effects and with an association between level-2 and level-3 class membership do not perform better than the baseline model, which indicates that there is no need to include these effects. However, the last model has lower BIC values than the baseline model, which indicates that there is evidence that the level-2 class intercept differences vary across level-3 classes. This model will serve as our final model.

{INSERT FIGURES 1a, 1b AND 1c ABOUT HERE}

Figure 1a depicts the estimated value of the intercept for each level-2 and level-3 class combination (these are obtained by adding up the fixed and the random intercept terms, including the interaction), Figure 1b the time effects per level-2 class (these sum to 0 across time points and are obtained by adding up the fixed and the random time effects), and Figure 1c the religion effects per level-3 class (these sum to 0 across religion categories and are obtained by adding up the fixed and the random religion effects). Figure 1a and 1b show that level-2 class one contains the respondents who are most against abortion, irrespective of the level-3 (region) class, and whose opinion is most stable across the four measurement occasions. Depending on the region class, class 2 respondents are very much in favor or somewhat against abortion, and they become less in favor during the observation period. Class 3 is (moderately) against, and shows a decrease and subsequently a return to the initial position during the four years of the study. Class 4 is (moderately) in favor of abortion, but much less at the first two occasions than at the last two occasions.

As far as the level-3 classes is concerned, Figure 1a shows that level-3 classes of regions differ in opinion concerning abortion only among class two respondents.

Moreover, the religion effects (Figure 1c) are small in class 2, with “other religion” slightly more against and “no religion” slightly more in favor of abortion. In class 1 regions, the religion effects are huge: here, Catholics are much more against abortion that the other religion categories. The latter ones show similar mutual differences as in the other latent class.

This example showed that complex but interesting level-2 and level-3 variability in intercepts and slopes can be detected using model specifications which are rather straightforward within the multilevel LC analysis framework. The most similar specification using a “standard” three-level logistic regression model would be a model with 4 normally distributed random effects at level-2 and 4 at level-3. Interpretation of the results of such an analysis would probably have been more difficult than the results presented above.

5.2 Multilevel mixture modeling with a set of continuous responses

The data for this example were collected by Van Peet (1992) and used by Hox (2002) to illustrate multilevel factor analysis (FA). Six continuous measures supposed to be connected to intelligence – “word list”, “cards”, “matrices”, “figures”, “animals”, and “occupations” – are available for 269 children belonging to 49 families. For 82 children, there is partially missing information, but these observations can be retained in the analysis using standard ML estimation methodology with missing data.

Hox (2002) analyzed this data set (excluding cases with missing values) using multilevel FA, which basically involves performing separate analyses of the within- and between-family covariance matrices. At the within level, his final model contained a “numeric” factor (loading on word list, cards, and matrices) and a “perception” factor (loading on figures, animals, and occupations), whereas at the between level a single factor sufficed. His aim was to determine whether the six measures relate to similar aspects of intelligence at the within- and between-family level, as well as to detect possible family effects, which may be explained by genetic and/or common environment factors.

To illustrate various types of multilevel mixture models, we will analyze this data set in three different ways, in each of which level-3 variation is modeled by assuming that families belong to a small number of level-3 classes. The first analysis uses a model corresponding to the first special case discussed in section 3.1; i.e., a model in which level-2 classes affect the item responses and level-3 classes the level-2 class memberships, but not directly the responses. Between-family differences in responses are thus explained by between-family differences in the likelihood of belonging to the child-level intelligence clusters. This is also the specification Vermunt (2008a) used in an earlier analysis of this data set.

The second analysis is conceptually similar to Hox’s analysis, but with discrete instead of continuous latent variables. In this analysis, both latent variables are assumed to affect the responses directly. The role a particular item plays in the clustering of children and the clustering of families may be fully different. Moreover, the clustering of children is conditional on the family clustering, which means that it is based on within-family differences which remain after taking into account the differences between family classes.

The third analysis uses continuous latent variables at level two. The model is a multilevel mixture factor model, a model with a mixture distribution at level three to capture between-family differences in the parameters of the child-level factor model. This can be seen as a kind of multiple group factor analysis with a large number of groups. The aim is to investigate whether the factor model parameters can be assumed to be invariant across groups.

Preliminary analysis showed that simple univariate normal within-class distributions with homogeneous residual variances across lower- and higher-level classes can be assumed for the six response variables. More specifically, inspection or pairwise residuals showed that there is no need to allow for within-class correlations across responses, and comparison of models with equal and unequal variances showed that it is correct to assume that residual variances are homogeneous across lower- and higher-level classes.

Analysis 1: indirect effect of family classes

In this first analysis, the six intelligence measures were used to cluster children into intelligence classes and it was investigated whether (classes of) families differ in the distribution of (their) children over these "intelligence" clusters. Table 3 provides the fit measures for the estimated multilevel LC models. As can be seen, a model with four child-level classes and three family-level classes performed best according to both BIC(2) and BIC(3).

{INSERT TABLE 3 ABOUT HERE}

{INSERT FIGURES 2a AND 2b ABOUT HERE}

Figures 2a displays the estimated $\lambda_{qc}^{(1,2)}$ parameters appearing in equation (6) for the model with *C*=4 and *D*=3. These parameters (which sum to 0 over classes) show that the means of the six intelligence indicators are nicely ordered across child-level classes 1 to 3. These can therefore be labeled high, middle, and low. Children in class 4 show a somewhat mixed pattern: they perform better than the middle class on cards and figures, better than the low class on word list and matrices and worse than the low class on animals and occupations.

Figure 2b displays the estimated level-2 class membership probabilities for the level-3 classes. As can be seen, in family-level class 3 almost all children belong to the mixed child-level class. Children from families belonging to family-level class 1 are more likely to be in the high intelligence class and children from family-level class 2 are more often in the middle and low intelligence classes. These results show that there is a very strong family effect on the performance of children on these six intelligence subtests.

Analysis 2: direct effect of family classes

In this second analysis, the six intelligence measures are used to simultaneously find child-level and family-level intelligence classes based on the children’s responses. Table 4 provides the fit measures for the estimated multilevel LC models. As can be seen, the model with *C*=3 and *D*=3 performs best according to BIC(2), whereas the model with *C*=3 and *D*=4 performs best according to BIC(3). Because the discrepancy is in the number of level-3 classes, we decided to keep the model selected by BIC(3) as the final model. Note that the BIC values of this model are lower than the once found in the previous analysis, which indicates that the assumption we made earlier – that differences in responses across family classes are fully mediated by differences in child-level class membership – is not fully correct.

{INSERT TABLE 4 ABOUT HERE}

{INSERT FIGURES 3a AND 3b ABOUT HERE}

Figures 3a and 3b display the estimates for the $\lambda_{qc}^{(1,2)}$ and $\lambda_{qd}^{(1,3)}$ parameters from equation (6), which can be used to label the level-2 and level-3 classes. The parameters for the level-2 classes show that class 1 scores higher on all measures than classes 2 and 3. Class 2 scores higher than class 3 on the first four measures (with a large difference on cards), but lower than class 3 on animals and occupations. This pattern reveals that there is a kind of two-dimensional structure.

Although the fit measures indicated that there are significant differences between families, it is not easy to give a simple interpretation to the encountered differences between the level-3 classes. Contrary to the results by Hox, we do not find a one dimensional structure, which would imply that classes should be (almost) ordered. Class-1 families score high on all measures, except for occupations on which they have a medium level score. Families belonging to class 2 score high on figures, medium on cards, and low on the remaining four 4 items. Class 3 scores low on three items and medium on the remaining items. Class-4 families score extremely high on occupations, high on word list, figures and matrices, medium on cards, and somewhat low on animals.

Analysis 3: multilevel mixture factor analysis

In the third analysis, we used a factor analytic model at level-2. Similar to Hox's analysis, it is a two factor model, but with the difference that "figure" loads on both factors. Seemingly, the factor structure changes somewhat when retaining cases with missing values in the analysis. Our baseline multilevel mixture factor analysis model is a model in which factor (co)variances, means, and loadings, as well as item intercepts are allowed to differ across family classes. Measurement equivalence across families is achieved when factor loadings and item intercepts turn out to be the same across family clusters.

We fitted models with 1 to 4 family-level latent classes. Both BIC(2) and BIC(3) indicated that a model with 3 classes is the one that should be preferred. Restricting the factor (co)variances and loadings to be equal across classes did not deteriorate the fit of the model. However, assuming that also item intercepts are equal across classes yields a worse model fit. Actually, except for the two reference items (the two items for which intercepts were fixed to 0 to be able to identify the factor means), none of the item intercepts can be assumed to be equal across classes. This confirms the results we found in the models with a discrete level-2 latent variables, namely that it is not correct to assume that family effects on item responses can be assumed to be fully mediated by the child-level latent variable(s).

6. Discussion

Whereas typical applications of LC analysis concern single-level multivariate response data sets, in this chapter, I demonstrated how LC and mixture models can be used for analyzing univariate two-level data sets, univariate three-level data sets, and multivariate two-level data sets. I also discussed how multilevel LC models fit into a more general latent variable modeling framework, which allows defining models with discrete and continuous latent variables at the multiple levels of a hierarchical structure.

The multilevel LC models were illustrated using two empirical examples. The first example showed how to use a multilevel mixture models for three-level regression analysis. Complex but interesting level-2 and level-3 variability in intercepts and slopes were detected using model specification which are rather straightforward with the presented framework.

A second empirical data set was analyzed in three different ways. Which of the three is most appropriate depends on the exact aim of the research concerned. It is, of course, also possibly that none of the three is appropriate that another type of analysis should be used, e.g., the multilevel factor analysis used by Hox (2002). Our second analysis, as well as Hox's multilevel factor model are more suited for exploration, whereas our first and third analysis are more suited when the items can be assumed to be meaningful indicators for clustering or measuring one or more underlying factors at the lower level.

Other types of illustrations of multilevel mixture models than the ones presented here can be found in the literature. Vermunt (2003, 2005, 2007, and 2008a) gave examples of multilevel variants of standard LC models for categorical response variables. A type of model that was not illustrated with an example is the model containing continuous random effects at level 3 discussed in Section 3.2. Applications of this model are provided by Vermunt (2003, 2005). Other applications, which similarly to our third analysis of the intelligence data, use continuous variables at level 2 and discrete latent variables at level 3 are the multilevel mixture growth models proposed by Palardy and Vermunt (2008), the multilevel mixture IRT models used by Vermunt (2008b), and the multilevel mixture factor models by Varriale (2008).

Since multilevel mixture modeling is a rather new area statistical methodology, it is not surprising that many issues have not yet been fully resolved. Future research should deal with issues such as sample size requirements, model selection strategies, model diagnostics, and effects of model misspecification. Possible extensions of the models presented in this chapter are multilevel LC models with ordinal latent variables and multilevel variants of latent Markov models.

Appendix

The models in Table 2 can be estimated with either the Latent GOLD Regression or Syntax module. To illustrate the Latent GOLD Syntax language, these are possible "`variables`" and "`equations`" sections of a setup for the baseline models appearing in Table 2:

```
variables
   groupid District_ID;
   caseid Case_ID;
   dependent Response binomial exposure=7;
   independent Year nominal, Religion nominal;
   latent U3 group nominal 2, U2 nominal 4;
equations
   Response <- 1 + Year + Religion + U2 + U2 Year
               + U3 + U3 Religion;
   U2 <- 1;
   U3 <- 1;
```

The "`variables`" section defines the level-3 (group) and level-2 (case) identifiers, the dependent variable, the predictors (independent), and the latent variables in the model. Note that both latent variables are nominal, with 2 and 4 categories, respectively. Moreover, for `U3`, the keyword "`group`" indicates that it is a level-3 latent variable. The "`equations`" section contains, in fact, the regression equations (6)-(8), where "`1`" defines an intercept term and where dummies/effects are automatically set up for nominal variables.

The two restricted models appearing in Table 2 can be obtained by eliminating either "+ `U2 Year`" or "+ `U3 Religion`" from the model for the response variable. The more extended models are obtained by adding "+ `U3 Year`" to the model for the response variable, "+ `U3`" to the model for the level-2 classes, and "+ `U2 U3`" to the model for the response variable.

The setup used for the second example differs from the one above in that the model is defined as a two-level model for multivariate responses. In other words, we have a model for six dependent variables rather than for one, but no case identifier needs to be specified since we have only one record per case. Another difference is that "equations" need to be specified for the residual variances of the response variables, because these are assumed to be normally distributed. The setup used for the models in Table 3 is:

```
variables
   groupid family;
   dependent wordlist continuous, cards continuous, figures
      continuous, matrices continuous, animals continuous,
      occupations continuous;
   latent U3 group nominal 4, U2 nominal 3;
equations
   cards <- 1 + U2;
   figures <- 1 + U2;
   matrices <- 1 + U2;
   animals <- 1 + U2;
   occupations <- 1 + U2;
   wordlist;
   cards;
   figures;
   matrices;
   animals;
   occupations;
   U2 <- 1 + U3;
   U3 <- 1;
```

A more compact specification of the equations using variable lists is

```
equations
   cards - occupations <- 1 + U2;
   cards - occupations;
   U2 <- 1 + U3;
   U3 <- 1;
```

For the second analysis (models in Table 4), we remove “+ `U3`” from model for “`U2`” and include it in the equations for the dependent variables. The multilevel mixture factor model can be defined as follows:

```
variables
   groupid ...
   dependent ...
   latent U3 nominal group 3, F1 continuous, F2 continuous;
equations
   wordlist <- (1) F1;
   cards <- 1 | U3 + F1 | U3;
   figures <- 1 | U3 + F1 | U3 + F2 | U3;
   matrices <- 1 | U3 + F1 | U3;
   animals <- (1) F2;
   occupations <- 1 | U3 + F2 | U3;
   wordlist - occupations;
   F1 <- 1 | U3;
   F2 <- 1 | U3;
   F1 | U3;
   F2 | U3;
   F1 <-> F2 | U3;
   U3 <- 1;
```

This setup illustrates various additional syntax options: two continuous latent variables are defined in “`latent`”, “equations” are included for the factor means and (co)variances, the statement “`| U3`” is used to indicate that a parameter varies across level-3 clusters, and “`(1)`” is used to fix the parameter concerned to 1.

Table 1: Nine-fold classification of latent variable models for three-level data sets

Level-2 latent variables $U_{jk}^{(2)}$	Level-3 latent variables $U_{k}^{(3)}$		
	Continuous	Discrete	Combination
Continuous	A1	A2	A3
Discrete	B1	B2	B3
Combination	C1	C2	C3

Table 2. Fit measures for the models estimated with the abortion data

Model	LL	Npar	BIC(2)	BIC(3)
C=1 D=1 baseline	-2188	7	4416	4405
C=2 D=1 baseline	-1745	12	3558	3539
C=3 D=1 baseline	-1683	17	3460	3433
C=4 D=1 baseline	-1657	22	3436	3401
C=5 D=1 baseline	-1645	27	3441	3398
C=1 D=2 baseline	-2073	12	4212	4193
C=2 D=2 baseline	-1712	17	3519	3492
C=3 D=2 baseline	-1671	22	3465	3431
C=4 D=2 baseline	-1644	27	**3438**	**3396**
C=5 D=2 baseline	-1636	32	3450	3399
C=1 D=3 baseline	-1999	17	4092	4065
C=2 D=3 baseline	-1699	22	3520	3485
C=3 D=3 baseline	-1663	27	3477	3434
C=4 D=3 baseline	-1638	32	3455	3404
C=5 D=3 baseline	-1629	37	3465	3406
C=4 D=2 religion not random	-1651	24	3435	3397
C=4 D=2 time not random	-1683	18	3467	3439
C=4 D=2 time also random at level-3	-1639	30	3444	3397
C=4 D=2 association between U2 and U3	-1640	30	3448	3400
C=4 D=2 interaction between U2 and U3	-1632	30	**3432**	**3384**

Table 3. Fit measures for the models estimated with the intelligence data (first analysis)

Model	LL	Npar	BIC(2)	BIC(3)
C=1 D=1	-4238	12	8543	8522
C=2 D=1	-4149	19	8404	8372
C=3 D=1	-4127	26	8400	8356
C=4 D=1	-4113	33	8411	8355
C=5 D=1	-4104	40	8432	8364
C=2 D=2	-4130	21	8378	8342
C=3 D=2	-4108	29	8379	8330
C=4 D=2	-4087	37	8381	8318
C=5 D=2	-4075	45	8402	8326
C=2 D=3	-4130	23	8388	8349
C=3 D=3	-4098	32	8374	8320
C=4 D=3	-4072	41	**8374**	**8304**
C=5 D=3	-4060	50	8400	8315
C=2 D=4	-4130	25	8399	8357
C=3 D=4	-4096	35	8388	8328
C=4 D=4	-4070	45	8391	8315
C=5 D=4	-4052	55	8412	8318
C=2 D=5	-4130	27	8410	8364
C=3 D=5	-4096	38	8405	8340
C=4 D=5	-4069	49	8412	8329
C=5 D=5	-4050	60	8436	8334

Table 4. Fit measures for the models estimated with the intelligence data (second analysis)

Model	LL	Npar	BIC(2)	BIC(3)
C=1 D=1	-4238	12	8543	8522
C=2 D=1	-4149	19	8404	8372
C=3 D=1	-4127	26	8400	8356
C=4 D=1	-4113	33	8411	8355
C=5 D=1	-4104	40	8432	8364
C=1 D=2	-4155	19	8417	8385
C=2 D=2	-4104	26	8354	8310
C=3 D=2	-4086	33	8356	8300
C=4 D=2	-4070	40	8364	8296
C=5 D=2	-4061	47	8385	8305
C=1 D=3	-4132	26	8410	8366
C=2 D=3	-4081	33	8346	8290
C=3 D=3	-4059	40	**8342**	8274
C=4 D=3	-4045	47	8354	8274
C=5 D=3	-4034	54	8370	8278
C=1 D=4	-4114	33	8413	8357
C=2 D=4	-4068	40	8361	8293
C=3 D=4	-4045	47	8352	**8272**
C=4 D=4	-4033	54	8368	8276
C=5 D=4	-4025	61	8391	8287
C=1 D=5	-4101	40	8425	8357
C=2 D=5	-4058	47	8379	8299
C=3 D=5	-4036	54	8374	8282
C=4 D=5	-4021	61	8384	8280
C=5 D=5	-4015	68	8410	8294

Figure 1a. Intercept for all combinations of level-2 and level-3 classes obtained with abortion data

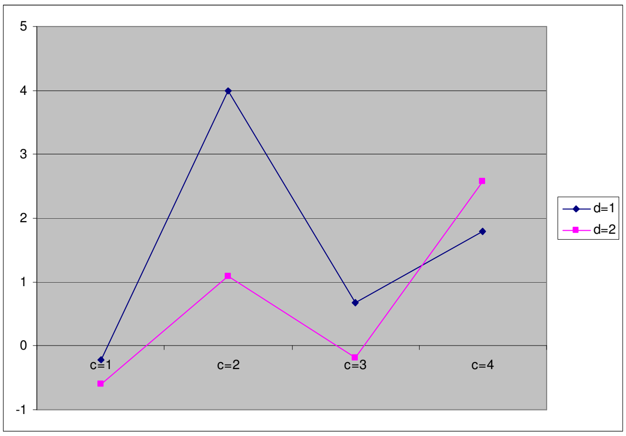


Figure 1b. Time effects for level-2 classes obtained with abortion data

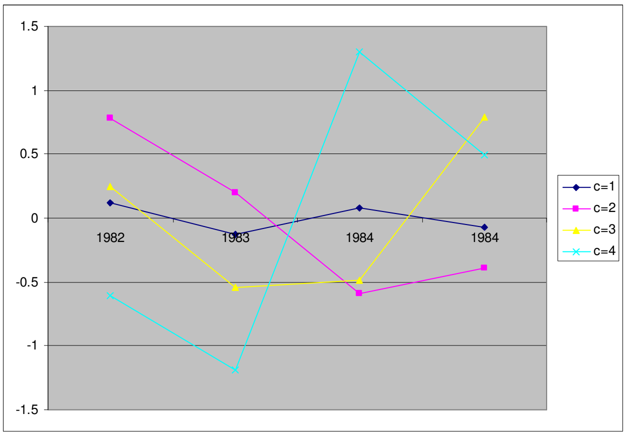


Figure 1c. Religion effects for level-3 classes obtained with abortion data

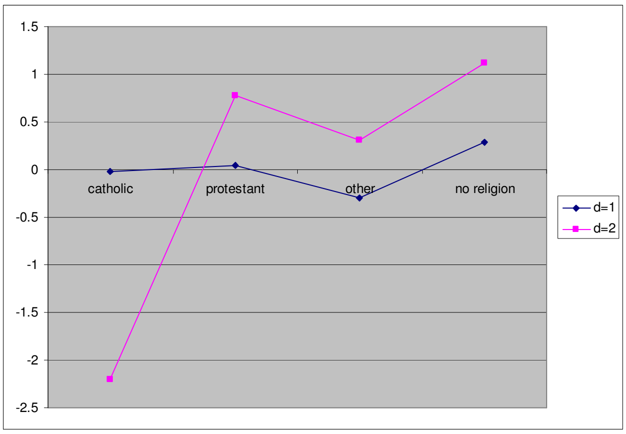


Figure 2a. Intercept differences between child-level classes obtained with intelligence data (first analysis)

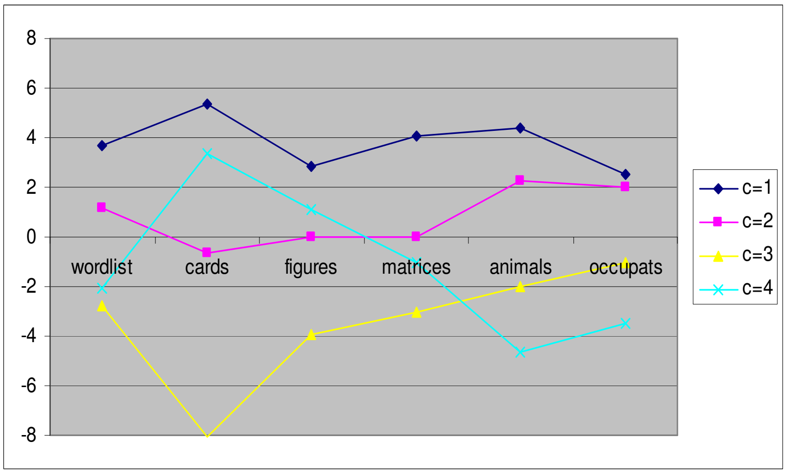


Figure 2b. Child-level class proportions for family-level classes obtained with intelligence data (first analysis)

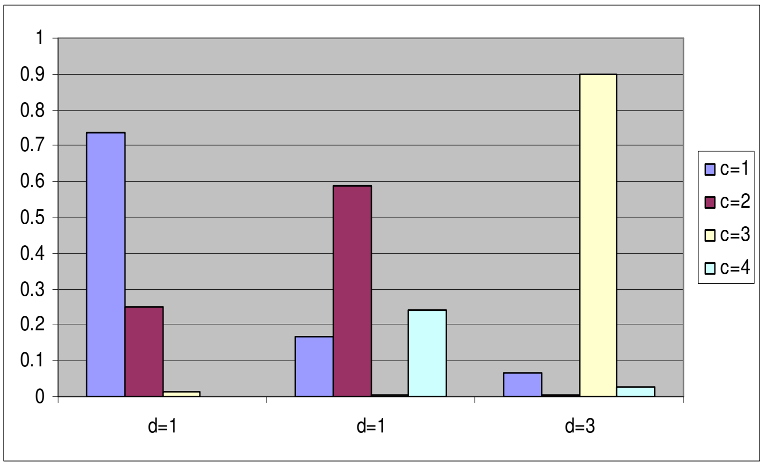


Figure 3a. Intercept differences between child-level classes obtained with intelligence data (second analysis)

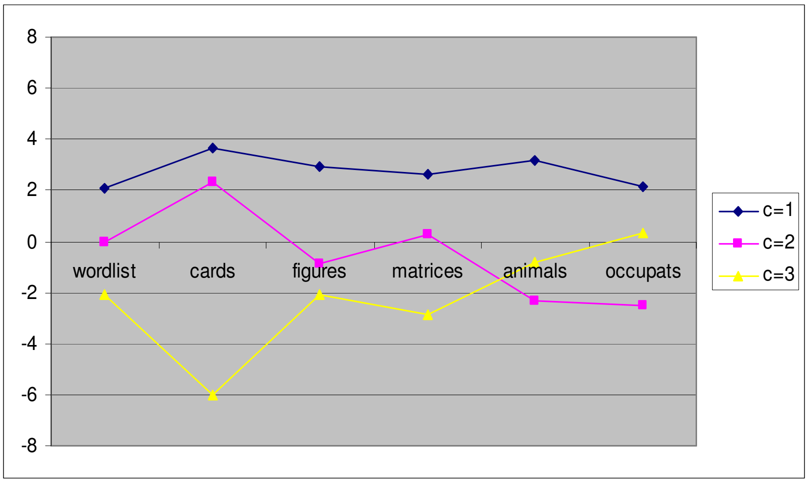
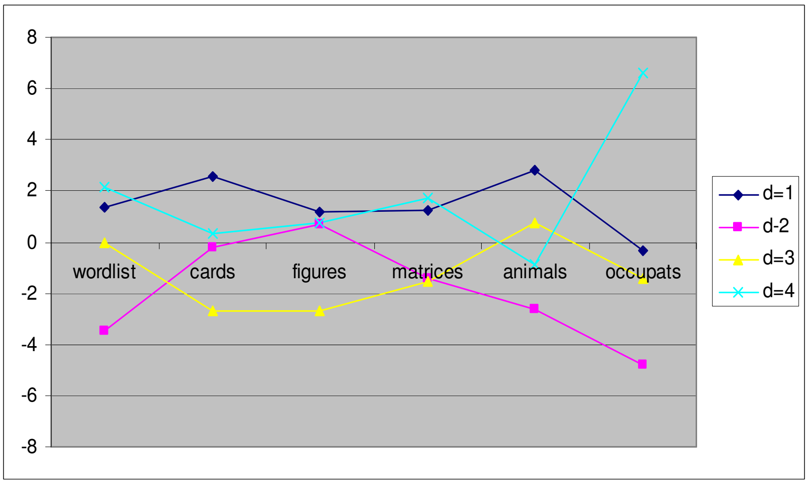


Figure 3b. Intercept differences between family-level classes obtained with intelligence data (second analysis)



References

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55, 218–234.

Allua, S.S. (2007). *Evaluation of Single- and Multilevel Factor Mixture Model Estimation*. Phd thesis, The University of Texas at Austin.

Asparouhov, T. & Muthen, B. (2008). Multilevel mixture models. G.R. Hancock and K.M. Samuelsen (eds.), *Advances in latent variable mixture models*, pp. 27-51. Charlotte, NC: Information Age Publishing, Inc.

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

Bouwmeester, S., Vermunt, J.K., and Sijtsma, K. (2007). Development and individual differences in transitive reasoning: a fuzzy trace theory approach. *Developmental Review*, 27, 41-74.

De Boeck, P., and Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.

Fox, J. P., and Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. Psychometrika, 66, 269-286.

Goldstein, H. (2003). *Multilevel statistical models* (3nd edition), London: Arnold.

Goldstein, H., and Browne, W. (2002). Multilevel factor analysis modelling using Markov chain Monte Carlo estimation. G.A. Marcoulides and I. Moustaki (eds.), *Latent variable and latent structure models*, 225-243. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Grilli L., and Rampichini C. (2007) Multilevel factor models for ordinal variables. *Structural Equation Modeling*, 14, 1-25.

Hagenaars, J.A. (1993). *Loglinear models with latent variables*. Newbury Park: Sage.

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, 22, 1433–1446.

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

Hox J. (2002). *Multilevel analysis: techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Kalmijn, M., and Vermunt, J.K. (2007). Homogeneity of social networks by age and marital status: A multilevel analysis of ego-centered networks. *Social Networks*, 29, 25-43.

Lazarsfeld, P.F., and Henry, N.W. (1968). *Latent Structure Analysis.* Boston: Houghton Mill.

Lukociene, O., Vermunt, J.K (2008). Determining the number of components in mixture models for hierarchical data. In: *Studies in Classification, Data Analysis, and Knowledge Organization*. Springer: Berlin-Heidelberg. In press.

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

Muthén, B.O. (1994) Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376-398.

Muthén, B. (2004). Latent variable analysis. Growth mixture modeling and related techniques for longitudinal data. D. Kaplan (ed.), *The Sage Hhandbook of quantitative methodology for the social sciences*, 345-368. Thousand Oakes: Sage Publications.

Muthén, B., and Muthén, L., (2008). *Mplus: User's manual*. Los Angeles: Muthén & Muthén.

Nagin, D.S. (1999). Analyzing developmental trajectories: a semiparametric group-based approach. *Psychological Methods*, 4, 139-157.

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

McGrath, K., and Waterton, J. (1986). *British social attitudes, 1983-1986 panel survey*. London: Social and Community Planning Research, Technical Report.

Palardy, G., and Vermunt, J. K. (2008). *Multilevel growth mixture models for classifying group-level observations*, submitted.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika*, 69, 183-206.

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

Van de Pol, F., and Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, 20, 213-247.

Van Peet A.A.J. (1992). *De potentieeltheorie van intelligentie* (The potentiality theory of intelligence). University of Amsterdam, Phd thesis.

Varriale, R. (2008). *Multilevel mixture models for the analysis of the university effectiveness.* Phd thesis, University of Florence.

Vermunt, J.K. (1997). *Log-linear models for event histories*. Thousand Oakes: Sage Publications.

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. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica*, 58, 220-233.

Vermunt J.K (2005). Mixed-effects logistic regression models for indirectly observed outcome variables. *Multivariate Behavioral Research*,40, 281-301.

Vermunt, J.K. (2007a. A hierarchical mixture model for clustering three-way data sets. *Computational Statistics and Data Analysis*, 51, 5368-5376.

Vermunt, J.K. (2007b). Growth models for categorical response variables: standard, latent-class, and hybrid approaches. K. van Montfort, H. Oud, and A. Satorra (eds.), *Longitudinal models in the behavioral and related sciences*, 139-158. Mahwah, NJ: Erlbaum.

Vermunt, J.K. (2008a. Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, 17, 33-51.

Vermunt, J.K. (2008b). Multilevel latent variable modeling: An application in educational testing, *Austrian Journal of Statistics*, 37, 285–299.

Vermunt, J.K. Langeheine, R., and Böckenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, 178-205.

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

Vermunt J. K., and Magidson J. (2005) *Latent GOLD 4.0 user's guide*. Belmont, MA: Statistical Innovations Inc..

Vermunt, J. K. and Magidson, J. (2008). *LG-syntax user's guide: Manual for Latent GOLD 4.5 syntax module*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K., Tran, B., and Magidson, J. (2008). Latent class models in longitudinal research. S. Menard (ed.), *Handbook of longitudinal research: Design, measurement, and analysis*, pp. 373-385. Burlington, MA: Elsevier.

Vermunt, J.K. and Van Dijk. L. (2001). A nonparametric random-coefficients approach: the latent class regression model. *Multilevel Modelling Newsletter*, 13, 6-13.

Wedel, M., and DeSarbo, W. (1994). A review of recent developments in latent class regression models. R.P. Bagozzi (ed.), *Advanced methods of marketing research*, 352-388. Cambridge, MA: Blackwell Publishers.

Yung, Y.F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62, 297-330.