Conditional Two Level Mixture with Known Mixing Proportions: Applications to School and Student Level Overweight and Obesity Data from Birmingham, England

Shakir Hussain¹,*, Mehdi AL-Alak² and Ghazi Shukur³

¹School of Health and Population Science, University of Birmingham, Birmingham, UK
²Central Organization for Statistics, Baghdad, Iraq
³Department of Economic and Statistics, Linnaeus University, Sweden

Abstract: Two Level (TL) models allow the total variation in the outcome to be decomposed as level one and level two or 'individual and group' variance components. Two Level Mixture (TLM) models can be used to explore unobserved heterogeneity that represents different qualitative relationships in the outcome.

In this paper, we extend the standard TL model by introducing constraints to guide the TLM algorithm towards a more appropriate data partitioning. Our constraints-based methods combine the mixing proportions estimated by parametric Expectation Maximization (EM) of the outcome and the random component from the TL model. This forms new two level mixing conditional (TLMc) approach by means of prior information. The new framework advantages are: 1. avoiding trial and error tactic used by TLM for choosing the best BIC (Bayesian Information Criterion), 2. permitting meaningful parameter estimates for distinct classes in the coefficient space and finally 3. allowing smaller residual variances. We show the benefit of our method using overweight and obesity from Body Mass Index (BMI) for students in year 6. We apply these methods on hierarchical BMI data to estimate student multiple deprivation and school Club effects.

Keywords: Parametric Expectation Maximization, Multilevel Mixture, Conditional Multilevel Mixture Known Mix, Overweight and Obesity Data.

1. INTRODUCTION

In this paper we present methodology which is related to the common statistical method for the analysis of data with heterogeneous outcomes in nested and non-nested structures, for example, see [1]. Traditional approaches fail to appropriately estimate the mixed pattern in the data. Alternatively, the mixture model, overcomes these lacks and delivers appropriate results.

The term two level mixture (TLM) can have two meanings. The first refers to latent class models in which the probability of class membership is predicted by some covariates, i.e. the class membership is a function of the predictors. However, in other contexts, TLM is also used to refer to some specified number of latent classes as part of estimating the regression model; that is, the class membership is a function of the covariates see [2]. Wedel & DeSarbo (2002) [3] categorize the outcome and run separate regression on each class. Muthén & Asparouhov (2009) [2] developed an efficient TLM Model that is more general and Vermunt (2008) [4] illustrated TLM with three applications.

We focus on exploring outcome heterogeneity using the parametric Expectation Maximization (EM) method. Dempster et al. (1977) [5] have proposed these methods for the exponential family. In order to maximize the latent class model log-likelihood function the EM algorithm is used. Our methods rely on user-provided constraints to guide the TLM algorithm towards a more appropriate data partitioning. We assume that some pre-existing knowledge about the desired partitioning is available and we provide this knowledge in the form of constraints. The conditional two level mixture (TLMc) model is implemented in two stages; in the first stage we generated class indicators for each observation by using parametric EM and, in the second stage, we ran the hierarchical multilevel using prior information provided by the indicator (labels) generated in the first stage.

The Bayesian Information Criterion (BIC) that Schwarz (1978) [6] used for TLM attains optimality by trial and error. For example, we may run the marginal model three times: the first run includes two latent classes; then we add another latent and examine the BIC; and finally we use four latent classes and again examine the BIC. The BIC concludes that three latent classes are optimum, see [2].

It is well known that overweight or obesity represent the upper class of the continuous Body Mass Index

*Address correspondence to this author at the Department of Economic and Statistics, Linnaeus University, Sweden; Tel: +46 (0)470 708777; Fax: +46 (0)479 82478; E-mail: shakir.hussain@lnu.se, shakir_stat@hotmail.com
two level mixture model and its likelihood. In Section 3, we consider a two level mixture model and its likelihood. We present and illustrate our method using real dataset. Finally, we give a brief summary and conclusion in Section 5.

In our BMI example, an EM algorithm was used to estimate the distinct number of latent class and the mixture proportion for the estimated classes before parameter estimation of TLMc. However, the number of latent classes and the mixture proportion for TLM were estimated together at the same stage. This means that the two methods use different class memberships and we have different membership probability. We observe these hugely discrepant results between the two methods, with the TLM method showing huge group and individual residual variance compared to TLMc. Unlike the TLM model, the conditional TLMc offers medically meaningful coefficient space at student and school levels and these provide reasons to consider TLMc.

Our new approach is to build a student level screening device to uncover distributions within mixture models. Hence we shall be looking for a hidden latent class first, which could manifest itself as a heterogeneous element of the data. Since the preliminary examination of the whole data distribution could not show heterogeneity, any prior assumptions will not be made with regard to the nature of the possible heterogeneity. A mixture of normal distributions to the BMI is attempted to be fit as a mixture of numerous normal variables using the parametric EM method. Then, using the TL, TLM and TLMc, covariate association with BMI at individual school levels and these provide reasons to consider TLMc.

The rest of the paper is organized as follows:

Section 2 discusses the normal mixture model, its likelihood and its prior structure. In Section 3, we consider a two level mixture model and its likelihood. We present and illustrate our method using real dataset in Section 4. Finally, we give a brief summary and conclusion in Section 5.

2. ESTIMATING THE MIXTURE PROPORTIONS BY EM ALGORITHM

In clustering analysis, mixture probability densities are commonly used; the standard algorithm for learning clusters from the data is the EM method. Searching for and identifying clusters can be used in the classification of a new data point or for predicting missing data. A useful and popular class of models is mixture model; see [8]. Typically the EM model components are Gaussian density function and assumed to be generated by hypothetical Gaussian mixture. Because of their probabilistic nature, Gaussian mixtures are in principle preferred over models that partition a data set into discrete parts. In most applications where a new data item needs to be classified, it is more desirable to calculate the probability that this item belongs to a certain cluster than to assign it to strictly one specific cluster. We will use the parametric EM method to capture the unobserved heterogeneity of possible classes in Gaussian mixture measurement by estimating the mixture proportions in the form of categorical latent variables and estimate the first two moments in each class assuming bimodal distribution.

Let the random variable S be a mixture of several m \( \geq 2 \) normal distributions. That is \( S_i \sim N(\mu_i, \sigma_i) \) for \( i = 1, 2, \ldots, m \) then we may write

\[
S = \sum_{i=1}^{m} R_i . S_i ,
\]

where \( R_i \in \{0, 1\} \) with \( p(R_i = 1) = \pi_i \), such that \( \sum \pi_i = 1 \) and the joint distribution of the binary vector \( (R_1, \ldots, R_m) \) is multinomial. If we generate multinomial variables \( (R_1, \ldots, R_m) \) with \( p(R_i = 1) = \pi_i \) then the density of S is

\[
P(S | \theta) = \sum_{i=1}^{m} \pi_i p(S | \theta_i) ,
\]

where \( \pi_i \) is the mixing proportion and \( p(\cdot | \theta_i) \) is a normal density with the parameter \( \theta_i = (\mu_i, \sigma_i) \) for \( i = 1, 2, \ldots, m \).

Fraley & Raftery (2002) [9] use the mixture in Gaussian clustering with m components, where the following likelihood,

\[
L(\theta_1, \ldots, \theta_m; \pi_1, \ldots, \pi_m | s) = \prod_{i=1}^{m} \sum_{k=1}^{m} \pi_k f_k(s | \theta_i) ,
\]

has the density \( f_k \) and parameters \( \theta_i \) of the \( k_{th} \) component in the mixture \( \pi_k \), which is the probability.
that an observation belongs to the $k$th component

$(\pi_k \geq 0; \sum_{k=1}^{m} \pi_k = 1)$.

The data can be viewed as consisting of $n$ multivariate observations with $r_i$ recoverable part of the $(s_i, z_i)$ in which $s_i$ is observed and $z_i$ is unobserved. The likelihood

$L(r_i | \theta) = \prod_{i=1}^{n} f(r_i | \theta), \quad (4)$

is then maximized to obtain the estimate of $\theta$. If the probability of a particular variable is unobserved depends only on the observed data $s$ and not on $z$, then the observed data likelihood can be obtained by integrating $z$ out of the complete data likelihood,

$L(s | \theta) = \int L(r | \theta) dz. \quad (5)$

The EM algorithm alternates between two steps, an ‘E’ step and an ‘M’ step, see [9].

The EM mixture model considers the complete data set $r_i = (s_i, z_i)$, where $z_i = (z_{i1}, \ldots, z_{im})$ and is the unobserved portion of the data with

$z_{ik} = \begin{cases} 1 & \text{if } r_i \text{ belongs to group } k \\ 0 & \text{otherwise.} \end{cases}$

Assume that each $z_i$ is independently and identically distributed according to a multinomial distribution of one draw from $m$ categories with probabilities $(\pi_1, \pi_2, \ldots, \pi_m)$, and that the density of an observation $s_i$ given $z_i$ is given by $\prod_{k=1}^{m} f(s_i | \theta_k)^{z_{ik}}$, the resulting complete data loglikelihood is

$l(\theta, \pi_k, z_{ik}, x_i) = \sum_{i=1}^{n} \sum_{k=1}^{m} z_{ik} \log(\pi_k f_i(S_i | \theta_k))$.

In fact the coding of the allocation estimate in the E-step follows

$z_{ik} \leftarrow \frac{\hat{\pi}_k f_i(S_i | \theta_k)}{\sum_{k=1}^{m} \hat{\pi}_k f_i(S_i | \theta_k)} \quad (6)$

and the M-step involves maximizing the loglikelihood in terms of $\pi_k$ and $\theta_k$ with $z_{ik}$ fixed at the values computed in the E step, $z_{ik}^*$. For an extensive discussion of the available implementation of the EM method for a variety of different parametric mixture models, see [8].

### 3. Multilevel Mixture with Unknown Mixing

We start by writing a simple empty (no covariates) multilevel model

$y_{ij} = \hat{\beta} + \beta_{0j} + e_{ij}, \quad (7)$

where $y_{ij}$ denotes the outcome for the $i$th individual of level one in the $j$th group of level two, $\hat{\beta}$ represents the grand mean, $\beta_{0j}$ is a random variable representing ‘between-units’ variability and $e_{ij}$ is a random variable representing ‘within-units’ variability.

The distributions of the random variables are assumed to be

$\beta_{0j} \sim N(0, \sigma^2_{\beta}), \quad e_{ij} \sim N(0, \sigma^2_e), \quad (8)$

where $\sigma^2_{\beta}$ and $\sigma^2_e$ are the variances of the between items (level two) and within items (level one) effects, respectively.

The model in (7) may include some covariates $x_{ij}$ at level one

$y_{ij} = \hat{\beta} + \beta_{1j} x_{ij} + e_{ij}, \quad (9)$

and other covariates $x_{ij}$ at level two

$\beta_{0j} = \lambda_0 + \lambda_1 x_{ij} + e_{ij}, \quad (10)$

where $\lambda_0$ and $\lambda_1$ are group level intercept and slope coefficients and both $e_{ij}$ and $\lambda_0$ are normals with mean zero and variances $\sigma^2_{\beta}$ and $\sigma^2_e$.

With a heterogeneous population and when regression of $y_{ij}$ on $x_{ij}$ vary across some latent class variable $C$ with $m$ categories, the residual $\eta_{ij}$ in

$y_{ij}^{C} = \hat{\beta} + \beta_{00} + \beta_{1j} x_{ij} + \eta_{ij}, \quad (11)$

may distributed as $\eta_{ij} \sim N(0, \theta_c)$, with covariance matrix $\theta_c$ reflecting the heterogeneity of residual variances at individual level, intercept and slope at group level and there covariance.

The probability of being in a given latent class with respect to a base class may vary as a function of a two-level multinomial logistic regression. For level one the probability of class membership presented in 12 below
where \( \alpha_{ij} \) stands for the random intercepts and \( x_{ij} \) represent the covariates.

In level two we have the following three equations, the random intercept, the random slope of equation (12) and the random intercept of equation (12):

\[
\beta_{0c} = \lambda_{00c} + \lambda_{01c} \omega_{0j} + u_{0j} \\
\beta_{1c} = \lambda_{10c} + \lambda_{11c} \omega_{1j} + u_{1j} \\
\alpha_{cj} = \lambda_{20c} + \lambda_{21c} \omega_{2j} + u_{2cj},
\]

(13) \hspace{1cm} (14) \hspace{1cm} (15)

The covariates \( (\omega_{0j}, \omega_{1j}, \text{and} \omega_{2j}) \) at group level are independent of the residuals \( (u_{0j}, u_{1j} \text{and} u_{2cj}) \).

Since TLM considers (12) as part of the optimization progression, sample allocations will be one element of the likelihood maximization procedure. The sample membership will not be constant and we may have different class membership estimate in TLM and in TLMc.

If the EM algorithm shows that the outcome is homogeneous and does not comprise distinct sub-populations then one can use TL rather than TLM or TLMc.

4. INTRODUCING THE CONDITIONAL TLMc METHOD

The individual latent class generated by parametric EM algorithm presented in Section 2

\[
f(s) = \sum_{k=1}^{\infty} \pi_k f_k(s | \theta_k)\]

gives a fixed number of classes, \( m \), probability density function, \( f_k \), and \( \pi_k \) the probability that an observation comes from the \( k \)th mixture \( (\pi_k \in (0,1) \text{ and} \sum_{k=1}^{\infty} \pi_k = 1) \).

Equation (12) becomes constant and no estimation for \( C_i \) is carried out; this is highlighted in

\[
p(C_i = c | x_i) = \frac{\exp(a_{ij} + b_i x_{ij})}{\sum_{c=1}^{\infty} \exp(a_{ij} + b_i x_{ij})} = \frac{\pi_k f_k(s_i | \theta_k)}{\sum_{c=1}^{\infty} \pi_k f_k(s_i | \theta_k)}.
\]

where the right hand side was determined in the early stage in Section 2.

This provides labels (class membership for each observation) or constraints to guide the algorithm towards a more appropriate data partitioning. The conditional TLMc format for \( m \) known classes with probability \( (\pi_1, \pi_2, \ldots \pi_m) \) is:

\[
y_{ij} | c = \beta_{01} + \beta_{1j} x_{ij} + \eta_{ij} \text{ with probability } \pi_1 \\
y_{ij} | c = \beta_{02} + \beta_{2j} x_{ij} + \eta_{2j} \text{ with probability } \pi_2 \\
\ldots \ldots \\
y_{ij} | c = \beta_{0m} + \beta_{mj} x_{ij} + \eta_{mj} \text{ with probability } \pi_m
\]

(16)

and the format of the random intercepts for \( m \) classes is:

\[
\beta_{0j} = \gamma_{01} + \gamma_{1j} u_{1j} + \zeta_j \\
\beta_{0j} = \gamma_{02} + \gamma_{2j} u_{2j} + \zeta_j \\
\ldots \ldots \\
\beta_{0j} = \gamma_{0m} + \gamma_{mj} u_{mj} + \zeta_j
\]

4.1. TL, TLM and TLMc Models for Overweight Data

As application, we are interested in two questions: 1) which individual level factors significantly predict overweight and obesity, and to what extent do they explain the observed level 1 variance and; 2) do any of the measured school level physical activity variables explain the observed level 2 variance, and to what extent?

Year 6 data will be used to illustrate our method. Student level variable, IMD and school level variable and proportion of pupils who participated in a club (Club) will be considered. Here, a single covariate is used for simplicity of illustration, but further covariates can clearly be added. A total of 5566 students in 147 schools make individual and group level sample sizes. The data and the MPLUS code for TLM and TLMc models are available from the author upon request.

We fit a sequence of a mixture of normal distributions with increasing numbers of latent class using the parametric EM algorithm to BMI measurements for students in Year 6. A best model can be estimated by fitting with different parameterization and/or a different number of classes and then applying a statistical criterion for model selection. Table 1 gives the estimates and the BIC.
Note that the estimated means differ by four BMI units and the BIC is (-29072) for the two-latent class, m = 2 model.

For m = 3, the mean estimate differs by 3 and 4 with BIC reduced by (150). When m = 4, the mean estimate differs by (1.9, 2.6, 3.2) and BIC increased by (12). The three latent classes are distinguished by the level of the first and second moments.

Monotone increase in the variance (1.95, 4.45, 17.85), and maximum BIC is attained for three latent classes. The plot of the BIC in Figure 1 indicates a best fit with m = 3 classes. The above prior information will be used as a constraint to run the conditional TLMc model.

The parameters estimated by the EM method indicate heterogeneity of the data, which seems to consist of three different types of BMI. This is, obviously, related to a high volume of BMI in the interval (11.89 to 41.52) for class three 'overweight or obese', for class two 'moderate' and for class one 'normal'.

The box plot in Figure 2 shows the parameter estimate of the three mixtures. The three latent classes are ordered from low to high BMI: 16.55 (class 1, 39%), 19.56 (class 2, 41%) and 23.59 (class 3, 20%). The mean differences correspond to (3 to 4) standard deviations of the BMI. Significant number of outlier points appears at third latent class when BMI goes above 31.
The density estimate $z_{ik}$ in equation (6) for membership in latent class $k$ is presented in Figure 3; this density show a right skewed normal distribution and it allows complete data representation with 3 distinct latent classes.

Heavy tail density shows for class 3 ‘overweight’. This may lead to high residual variance. The class indicator generated by parametric EM will be used for the TLMc model next. The other two densities are not heavy tailed; they may come from proper normal distributions.

The overweight or obesity study raises two main questions: which individual level predictors predict being obese; and does any of the measured school level predictors explain being obese?

This study focuses on nested sources of variability: students nested within schools. Group and individual level predictors will be considered.

In the initial Bayesian TL model we wish to fit

$$y_i \sim N(\alpha_{ji} + \beta x_i, \sigma_y^2), \quad \text{for } i = 1, \ldots, 5566$$

$$\alpha_j \sim N(\gamma_0 + \gamma_j, \sigma^2), \quad \text{for } j = 1, \ldots, 147$$

where $(\sigma_y^2$ and $\sigma^2_{\alpha})$ are individual and group residual variances and the individual and group covariates are $(x_i$ and $u_j$), respectively.

The Bayesian approach to the TL model in 17 adds flexibility to standard modelling. Prior information allows results of a previous model to be used to inform the
current model. The Markov Chain Monte Carlo (MCMC) method uses exact estimation instead of approximate and Bayesian inference via MCMC allow more complicated models that frequentists are unable to estimate.

We follow Gelman & Hill (2007) [10] and consider the non-informative prior distribution as “reference models” to be used as a standard of comparison or starting point in place of the proper, informative prior distribution. Uniform distribution is given to individual and group standard deviation $\sigma_y$ and $\sigma_a$. A uniform distribution means that the posterior distribution has the same shape as the likelihood function, which in turn means that the resulting Bayesian intervals and estimate will essentially match the traditional results. The normal distribution assigned to $\beta$ and $\gamma_0$ and $\gamma_1$ can be thought of as prior distributions for these intercept and slope. The BUGS code to compute posterior estimates from the TL model is presented below.

```plaintext
model
{
  for (i in 1:n){
    BMI[i] ~ dnorm (y.hat[i], tau.y)
    y.hat[i] <- a[ID[i]] + b*IMD[i]
  }
  b ~ dnorm (0, .0001)
  tau.y <- pow(sigma.y, -2)
  sigma.y ~ dunif (0, 100)
  for (j in 1:J){
    a[j] ~ dnorm (a.hat[j], tau.a)
    a.hat[j] <- g.0 + g.1*Club[j]
  }
  g.0 ~ dnorm (0, .0001)
  g.1 ~ dnorm (0, .0001)
  tau.a <- pow(sigma.a, -2)
  sigma.a ~ dunif (0, 100)
}
```

Summary statistics provided by WinBUGS are presented in Table 2 below. For detailed discussion see [11].

The mean and the standard deviation (SD) are simply the empirical average and the standard deviation of the sampled values, the MC (Monte Carlo) error provide an assessment of the sampling error on the mean attributable to the number of iterations performed. The 2.5%, median, and 97.5% are the empirical percentiles, while start is the iteration at which monitoring began and sample indicates the total number of iterations contributing to the summary statistics.

The TL model recognizes that the student level covariate IMD has a positive significant effect on all students’ BMI. This suggests that a 100-unit difference in IMD causes almost a unit increase in student BMI. For group level covariates, Club is not significant. Interclass correlation of the two variances is $0.1764/(0.1764 + 13.1044) = 0.01328233$. This indicates that the variance accounted for by school level is 1.33%. This indicates that school contribution is minor.

Detection of deviance convergence is presented in Figure 4. The continuous line joining successive realisations of the deviance plotted against Gibbs iteration number. A Markov chain shows a random scatter about a stable mean value. Three chains with three different colours presented in Figure 4 shows the fluctuation of each from the common mean.

A summary of BUGS simulations for TL model is presented in Figure 5; R-hat is near 1 and below 1.5 for all parameters, indicating approximate convergence.

The box plots in the top right panel presents the estimate of the random intercepts. The other TL model coefficients and variances are presented below.

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>SD</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>18.85</td>
<td>0.141</td>
<td>0.0108</td>
<td>18.58</td>
<td>18.850</td>
<td>19.12</td>
<td>501</td>
<td>1500</td>
</tr>
<tr>
<td>IMD</td>
<td>0.0102</td>
<td>0.003</td>
<td>2.6E-4</td>
<td>0.005</td>
<td>0.0100</td>
<td>0.016</td>
<td>501</td>
<td>1500</td>
</tr>
<tr>
<td>Club</td>
<td>0.002</td>
<td>0.005</td>
<td>2.6E-4</td>
<td>-0.008</td>
<td>0.0017</td>
<td>0.011</td>
<td>501</td>
<td>1500</td>
</tr>
<tr>
<td>sigma.a</td>
<td>0.42</td>
<td>0.071</td>
<td>0.0059</td>
<td>0.273</td>
<td>0.4217</td>
<td>0.570</td>
<td>501</td>
<td>1500</td>
</tr>
<tr>
<td>sigma.y</td>
<td>3.62</td>
<td>0.036</td>
<td>8.9E-4</td>
<td>3.551</td>
<td>3.6200</td>
<td>3.687</td>
<td>501</td>
<td>1500</td>
</tr>
<tr>
<td>deviance</td>
<td>3012</td>
<td>16.26</td>
<td>1.0040</td>
<td>301.0</td>
<td>3012</td>
<td>302.0</td>
<td>501</td>
<td>1500</td>
</tr>
</tbody>
</table>
Figure 4: Deviance convergence from 3 chains.

Figure 5: Bayesian TL model summary.
In what follows we apply two models: the first is the marginal TLM due to [2]

\[
y_{ij} = \beta_{ij} + \beta_{1i} x_{ij} + \eta_{ij}
\]

\[
p(C_y = c | x_y) = \frac{\exp(a_{ij} + b x_{ij})}{\sum_{c=1}^{K} \exp(a_{ij} + b x_{ij})},
\]

(18)

and the second is our proposed conditional TLMc model

\[
y_{ij} = \beta_{0ij} + \beta_{1ij} x_{ij} + \eta_{ij} \quad \text{with probability } \pi_1 = 0.39
\]

\[
y_{ij} = \beta_{0ij} + \beta_{1ij} x_{ij} + \eta_{ij} \quad \text{with probability } \pi_2 = 0.41
\]

\[
y_{ij} = \beta_{0ij} + \beta_{1ij} x_{ij} + \eta_{ij} \quad \text{with probability } \pi_3 = 0.20
\]

(19)

The key result is that school level covariate ‘Club’ comes out negative and significant in latent class 2 (moderate level of BMI) and class 3 (Overweight), with the conditional TLMc model only. In other words, our research using TLMc would conclude that increasing school percentage of club participation causes a reduction in BMI for two distinct classes “Moderate” and “Overweight”, but not for “Normal”.

This means that there is a variation in the response between the latent classes. For the first latent class ‘Normal’, no effect of Club is seen on student BMI; the magnitude of significant effects for class 2 to 3 is 16:32. The effects double for overweight students.

It is interesting to note that the IMD and Club influence on BMI is different in the three latent classes. The latent classes are distinguished not only by the

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Marginal TLM model</th>
<th>Conditional TLMc model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student Level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latent 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept &amp;nbysp; &amp;nbysp; &amp;nbysp;</td>
<td>16.656 (0.144)          </td>
<td>16.267 (0.068)  </td>
</tr>
<tr>
<td>IMD score  </td>
<td>0.0001 (0.002)    </td>
<td>0.001 (0.001)    </td>
</tr>
<tr>
<td>Residual Variance</td>
<td>2.136 (0.142)</td>
<td>1.254 (0.039)</td>
</tr>
<tr>
<td>Latent 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept  </td>
<td>19.756 (0.281)    </td>
<td>20.087 (0.104)    </td>
</tr>
<tr>
<td>IMD score</td>
<td>0.013 (0.005)</td>
<td>0.006 (0.002)</td>
</tr>
<tr>
<td>Residual Variance</td>
<td>5.777 (0.456)</td>
<td>2.127 (0.046)</td>
</tr>
<tr>
<td>Latent 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept  </td>
<td>23.307 (0.668)    </td>
<td>25.982 (0.287)    </td>
</tr>
<tr>
<td>IMD score</td>
<td>0.041 (0.012)</td>
<td>0.010 (0.006)</td>
</tr>
<tr>
<td>Residual Variance</td>
<td>17.148 (1.357)</td>
<td>8.349 (0.939)</td>
</tr>
<tr>
<td><strong>School Level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latent 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Club</td>
<td>0.014 (0.008)</td>
<td>0.100</td>
</tr>
<tr>
<td>Latent 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Club</td>
<td>-0.014 (0.019)</td>
<td>0.457</td>
</tr>
<tr>
<td>Latent 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Club</td>
<td>-0.040 (0.033)</td>
<td>0.223</td>
</tr>
<tr>
<td>Residual Variance</td>
<td>0.069 (0.024)</td>
<td>0.004</td>
</tr>
</tbody>
</table>
level of the BMI in the first part of our analysis, but also by the strength of the relation with IMD, (0.001, 0.006, 0.010) and Club, (0.002, -0.016, -0.032) in Table 3.

In the first stage of our method, the EM algorithm can identify distinct classes of BMI and in the second stage the conditional TLM can discover association effects of covariates on each BMI latent class.

The individual residual variances will be used for comparison of fit of the marginal TLM and the conditional TLM models.

In Figure 6, the residual variances estimate says that the marginal TLM is double the conditional TLMc in magnitude. The residual variance in latent class 3 is more than double that in latent class 2 for TLM, and is four times greater in TLMc. The level one variation is much larger than the level two variations, indicating greater unobserved heterogeneity on the individual student level. This analysis shows that TLM outperforms TL and TLM models.

5. CONCLUDING REMARKS

This paper presents a constraints-based method for allowing information regarding mixing proportions to be used in multilevel mixture models. Our approach provides more flexibility than the standard multilevel mixture. This flexibility can offer medically meaningful BMI profile in group and individual coefficient space and allow smaller residual variances.

For the conditional TLMc method to achieve its full potential objectives, the parametric EM algorithm was used to guide the TLM towards appropriate outcome partitioning “conditioned on the priori assumption based on the result of model-based clustering”. The marginal TLM method requires the user to run the model under different number of latent classes and choose the best model using the BIC, where the number of mixing is based on the BIC.

We prefer the overall conclusions of the conditional TLMc analysis. Our proposed method clearly shows that the association between IMD score and Club with BMI varies between three medical latent classes. This feature is not accommodated in the TL model. Using known mixtures, the conditional TLMc model is able to detect significant differences with Club which are not detected by the marginal TLM.

ACKNOWLEDGEMENTS

The authors thank Peymane Adab (University of Birmingham) and Paul Aveyard (Oxford University) for providing the BMI data to apply our Conditional Two Level Mixture model.
REFERENCES


