Mixture of Latent Trait Analyzers for Model-Based Clustering of Categorical Data

Isabella Gollini[†] & Thomas Brendan Murphy[⋆]

[†]National Centre for Geocomputation, National University of Ireland Maynooth, Ireland

[⋆]School of Mathematical Sciences, University College Dublin, Ireland

Abstract

Model-based clustering methods for continuous data are well established and commonly used in a wide range of applications. However, model-based clustering methods for categorical data are less standard. Latent class analysis is a commonly used method for model-based clustering of binary data and/or categorical data, but due to an assumed local independence structure there may not be a correspondence between the estimated latent classes and groups in the population of interest. The mixture of latent trait analyzers model extends latent class analysis by assuming a model for the categorical response variables that depends on both a categorical latent class and a continuous latent trait variable; the discrete latent class accommodates group structure and the continuous latent trait accommodates dependence within these groups. Fitting the mixture of latent trait analyzers model is potentially difficult because the likelihood function involves an integral that cannot be evaluated analytically. We develop a variational approach for fitting the mixture of latent trait models and this provides an efficient model fitting strategy. The mixture of latent trait analyzers model is demonstrated on the analysis of data from the National Long Term Care Survey (NLTCS) and voting in the U.S. Congress. The model is shown to yield intuitive clustering results and it gives a much better fit than either latent class analysis or latent trait analysis alone.

1   Introduction

Model-based clustering methods are widely used because they offer a coherent strategy for clustering data where uncertainties can be appropriately quantified using probabilities. Many model-

based clustering methods are based on the finite Gaussian mixture model (eg. Celeux and Govaert, 1995; Fraley and Raftery, 2002; McNicholas and Murphy, 2008) and these methods have been successfully applied to the clustering of multivariate continuous measurement data; more recent extensions of model-based clustering for continuous data use finite mixtures of non-Gaussian models (eg. Lin et al., 2007; Karlis and Santourian, 2008; Lin, 2010; Andrews and McNicholas, 2012).

Categorical data arise in a wide range of applications including the social sciences, health sciences and marketing. Latent class analysis and latent trait analysis are two commonly used latent variable models for categorical data (eg. Bartholomew et al., 2011). In the latent class analysis model, the dependence in the data is explained by a categorical latent variable that identifies groups (or classes) within which the response variables are independent (also known as the local independence assumption). Latent class analysis is used widely for model-based clustering of categorical data, however, if the condition of independence within the groups is violated, then the latent class model will suggest more than the true number of groups and so the results are difficult to interpret and can be potentially misleading. The latent trait analysis model uses a continuous univariate or multivariate latent variable, called a latent trait, to model the dependence in categorical response variables. If data come from a heterogeneous source or multiple groups, then a single multi-dimensional latent trait may not be sufficient to model the data. For these reasons, the proposed mixture of latent trait analyzers (MLTA) model is developed for model-based clustering of categorical data where a categorical latent variable identifies groups within the data and a latent trait is being used to accommodate the dependence between outcome variables within each cluster.

In this paper, we focus on the particular case of binary data and we propose two different mixture of latent trait analyzers models for model-based clustering of binary data: a general model that supposes that the latent trait has a different effect in each group and the other more restrictive, but more parsimonious, model that supposes that the latent trait has the same effect in all groups. Thus, the proposed family of models (MLTA) is a categorical analogue of some parsimonious model families used for clustering continuous data (eg. Fraley and Raftery, 2002; McNicholas and Murphy, 2008; Andrews and McNicholas, 2012). The MLTA family of models is most like the PGMM family (McNicholas and Murphy, 2008) which is based on the mixture of factor analyzers model (Ghahramani and Hinton, 1997) however the MLTA model accommodates binary response variables instead of continuous variables. In addition, the model can be easily applied to nominal categorical data by coding the data in binary form. The MLTA model is a generalization of the mixture of Rasch models (Rost, 1990), and is a special case of multilevel

mixture item response models (Vermunt, 2007); these connections are explored in more detail in Section 3.3, after the MLTA model has been fully introduced.

Fitting the mixture of latent trait analyzers model is potentially difficult because the likelihood function involves an integral that cannot be evaluated analytically. However, we propose using a variational approximation of the likelihood as proposed by Tipping (1999) for model fitting purposes because it is extremely easy to implement and it converges quickly, so it can easily deal with high dimensional data and high dimensional latent traits.

The mixture of latent trait analyzers model is demonstrated through the analysis of two data sets: the National Long Term Care Survey (NLTCS) data set (Erosheva et al., 2007) and a U.S. Congressional Voting data set (Asuncion and Newman, 2007). The latent class analysis and the latent trait analysis models are not sufficient to summarize these two data sets. However, the mixture of latent trait analyzers model detects the presence of several classes within these data and the dependence structure within groups is explained by a latent trait.

The paper is organized as follows. Section 2.1 provides an introduction to latent class analysis. Section 2.2 provides an introduction to latent trait analysis with a description of three of the most common techniques to evaluate the likelihood in this model: the Gauss-Hermite quadrature approach (Section 2.2.1), Monte Carlo sampling (Section 2.2.2), and the variational approach (Section 2.2.3). In Section 3 we introduce the mixture of latent trait analyzers model and a parsimonious version of the model (Section 3.1). An interpretation of the model parameters is outlined in Section 3.2, a discussion of models related to the MLTA model is given in Section 3.3, and a variational approach for estimating the model parameters is proposed in Section 3.4. The issue of model identifiability is discussed in Section 3.5. In Section 4 an adjustment to the BIC (Schwarz, 1978) is proposed as a model selection criterion and Pearson's χ^2 test and the truncated sum of squares Pearson residual criterion (Erosheva et al., 2007) are used to assess model fit. Computational aspects of fitting the MLTA model are outlined in Section 5. The two applications are presented in Section 6.1 and Section 6.2. We conclude in Section 7 by discussing the model and the results of its application.

2   Latent Class & Trait Analysis

In this section, we give an overview of latent class and latent trait analysis which are the basis of the mixture of latent trait analyzers model. We also outline how the models are fitted in practice.

2.1 Latent Class Analysis

The latent class analysis (LCA) can be used to model a set of N observations $\mathbf{x}_1, \ldots, \mathbf{x}_N$ each involving M binary (categorical) variables. Each observation $\mathbf{x}_n = (x_{n1}, x_{n2}, \ldots, x_{nM})$, where x_{nm} records the value of binary variable m for observation n where $x_{nm} \in \{0, 1\}$.

LCA assumes that there is a latent categorical variable $\mathbf{z}_n = (z_{n1}, z_{n2}, \ldots, z_{nG})$ for which $z_{ng} = 1$ if observation n belongs to class g and $z_{ng} = 0$ otherwise. We assume that $\mathbf{z}_n \sim$ Multinomial$(1, (\eta_1, \eta_2, \ldots, \eta_G))$, where η_g is the prior probability that a randomly chosen individual is in the gth class ($\sum_{g=1}^{G} \eta_g = 1$ and $\eta_g \geq 0$ for all $g = 1, \ldots, G$). Furthermore, conditional on the latent class variable $z_{ng} = 1$, the response variables $\mathbf{x}_n = (x_{n1}, \ldots, x_{nM})$ are distributed as independent Bernoulli random variables with parameters $\pi_{g1}, \pi_{g2}, \ldots, \pi_{gM}$; therefore, π_{gm} is the conditional probability of observing $x_{nm} = 1$ if the observation is from class g (that is, $z_{ng} = 1$). Hence, the dependence between variables is explained by the latent class variable. LCA can be seen as a finite mixture model with mixing proportions η_g in which the component distributions are multivariate independent Bernoulli distributions. Consequently the log-likelihood function is given as,

$$\ell = \sum_{n=1}^{N} \log \left(\sum_{g=1}^{G} \eta_g \prod_{m=1}^{M} \pi_{gm}^{x_{nm}} (1 - \pi_{gm})^{1-x_{nm}} \right). \tag{1}$$

Bartholomew et al. (2011) describe how to estimate the parameters by maximum likelihood estimation via the expectation-maximisation (EM) algorithm (Dempster et al., 1977).

When LCA is used in a model-based clustering context, the inferred latent classes are assumed to correspond to clusters in the data; this may be a valid conclusion when the model assumptions are correct but may not be otherwise.

2.2 Latent Trait Analysis

Latent trait analysis (LTA) can also be used to model a set of N multivariate binary (categorical) observations. LTA assumes that there is a D dimensional continuous latent variable $\mathbf{y}_n$ ($n = 1, \ldots, N$) underlying the behavior of the M categorical response variables within each observation $\mathbf{x}_n$ (Bartholomew et al., 2011). The LTA model assumes that

$$p(\mathbf{x}_n) = \int p(\mathbf{x}_n|\mathbf{y}_n)p(\mathbf{y}_n)\, d\mathbf{y}_n \tag{2}$$

4

where the conditional distribution of $\mathbf{x}_n$ given $\mathbf{y}_n$ is

$$p(\mathbf{x}_n|\mathbf{y}_n) = \prod_{m=1}^{M} p(x_{nm}|\mathbf{y}_n) = \prod_{m=1}^{M} (\pi_m(\mathbf{y}_n))^{x_{nm}} (1 - \pi_m(\mathbf{y}_n))^{1-x_{nm}}, \tag{3}$$

and the response function is a logistic function

$$\pi_m(\mathbf{y}_n) = p(x_{nm} = 1|\mathbf{y}_n) = \frac{1}{1 + \exp[-(b_m + \mathbf{w}_m^T \mathbf{y}_n)]}, \quad 0 \le \pi_m(\mathbf{y}_n) \le 1$$

where b_m is the intercept parameter and $\mathbf{w}_m$ are the slope parameters in the logistic function. Thus, the conditional probability that $x_{nm} = 1$ given $\mathbf{y}_n$ is an increasing function of $b_m + \mathbf{w}_m^T \mathbf{y}_n$ and if $\mathbf{w}_m = \mathbf{0}$ we get a probability equal to $1/(1 + \exp(-b_m))$ independently of the value of $\mathbf{y}_n$. Finally, it is assumed that $\mathbf{y}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Thus, although the variables within $\mathbf{x}_n$ are conditionally independent given $\mathbf{y}_n$, the marginal distribution of $\mathbf{x}_n$ accommodates dependence between the variables.

Therefore, the log-likelihood is

$$\ell = \sum_{n=1}^{N} \log \left(\int p(\mathbf{x}_n|\mathbf{y}_n) p(\mathbf{y}_n) \, d\mathbf{y}_n \right) = \sum_{n=1}^{N} \log \left(\int \prod_{m=1}^{M} p(x_{nm}|\mathbf{y}_n) p(\mathbf{y}_n) \, d\mathbf{y}_n \right) \tag{4}$$

The integral on (4) cannot be evaluated analytically, so it is necessary to use other methods to evaluate it. These are briefly reviewed in Section 2.2.1-Section 2.2.3.

2.2.1 Gauss-Hermite Quadrature

Bock and Aitkin (1981) proposed a method to evaluate the integral in (4) by using the Gauss-Hermite quadrature (Abramowitz and Stegun, 1964):

$$\ell \approx \sum_{n=1}^{N} \log \left(\sum_{q=1}^{Q} p(\mathbf{x}_n|\mathbf{y}_q) h(\mathbf{y}_q) \right) = \sum_{n=1}^{N} \log \left(\sum_{q=1}^{Q} \prod_{m=1}^{M} p(x_{nm}|\mathbf{y}_q) h(\mathbf{y}_q) \right),$$

where Q is the number of set of integration points and $h(\mathbf{y}_1), \ldots, h(\mathbf{y}_Q)$ ($\sum_{q=1}^{Q} h(\mathbf{y}_q) = 1$) are the weights for the sets of points $\mathbf{y}_1, \ldots, \mathbf{y}_Q$. In practice this method treats the latent variables as discrete taking values $\mathbf{y}_1, \ldots, \mathbf{y}_Q$ with probabilities $h(\mathbf{y}_1), \ldots, h(\mathbf{y}_Q)$. Thus, the model density can be approximated by a finite mixture model with Q component densities, where $p(\mathbf{x}_n|\mathbf{y}_1), \ldots, p(\mathbf{x}_n|\mathbf{y}_Q)$ are the component distributions and $h(\mathbf{y}_1) \ldots, h(\mathbf{y}_Q)$ are fixed mixing proportions. Bartholomew et al. (2011) explain this approach in detail and outline its drawbacks. The number of component densities, Q, required increases exponentially in D, so the Gauss-Hermite quadrature can be hard to implement and can be quite slow. Furthermore when the $\mathbf{w}_m$ parameters are very unequal their estimates can diverge to infinity.

2.2.2 Monte Carlo Sampling

An alternative approach is the Monte Carlo method (Sammel et al., 1997), that samples $\mathbf{y}_l$ ($l = 1, \ldots, L$) from the D-dimensional latent distribution, and approximates the log-likelihood as

$$\ell \approx \sum_{n=1}^{N} \log\left(\frac{1}{L}\sum_{l=1}^{L} p(\mathbf{x}_n|\mathbf{y}_l)\right) = \sum_{n=1}^{N} \log\left(\frac{1}{L}\sum_{l=1}^{L}\prod_{m=1}^{M} p(x_{nm}|\mathbf{y}_l)\right).$$

This approximation of the model density can be seen as a finite mixture model with component densities $p(\mathbf{x}_n|\mathbf{y}_1), \ldots, p(\mathbf{x}_n|\mathbf{y}_L)$ and the mixing proportions are equal to $\frac{1}{L}$, but usually L is quite large, making the implementation of this method quite difficult.

2.2.3 Variational Approximation

Tipping (1999) proposed to use a variational approximation (Jaakkola and Jordan, 1996) to fit the latent trait analysis model in a manner that is fast in convergence and easy to implement. The main aim of this approach is to maximize a lower bound that approximates the likelihood function. In latent trait analysis the log-likelihood function (4) is governed by the logistic sigmoid function, that can be approximated by the exponential of a quadratic form involving variational parameters $\boldsymbol{\xi}_n = (\xi_{n1}, \ldots, \xi_{nM})$ where $\xi_{nm} \neq 0$ for all $m = 1, \ldots, M$. This allows for the computation of an approximate log-likelihood in closed form. In this case the lower bound of each term in the log-likelihood is given by,

$$\mathcal{L}(\boldsymbol{\xi}_n) = \log\left(\tilde{p}\left(\mathbf{x}_n|\boldsymbol{\xi}_n\right)\right)$$
$$= \log\left(\int \prod_{m=1}^{M}\tilde{p}(x_{nm}|\mathbf{y}_n, \xi_{nm})p(\mathbf{y}_n)\,d\mathbf{y}_n\right),$$

where

$$\tilde{p}\left(x_{nm}|\mathbf{y}_n, \xi_{nm}\right) = \sigma(\xi_{nm})\exp\left(\frac{A_{nm}-\xi_{nm}}{2} + \lambda(\xi_{nm})\left(A_{nm}^2 - \xi_{nm}^2\right)\right),$$

$\sigma(\xi_{nm}) = (1 + \exp(-\xi_{nm}))^{-1}$, $A_{nm} = (2x_{nm}-1)\left(b_m + \mathbf{w}_m^T\mathbf{y}_n\right)$ and $\lambda(\xi_{nm}) = (\frac{1}{2} - \sigma(\xi_{nm}))/2\xi_{nm}$. This approximation has the property that $\tilde{p}\left(x_{nm}|\mathbf{y}_n, \xi_{nm}\right) \leq p\left(x_{nm}|\mathbf{y}_n\right)$ with equality when $|\xi_{nm}| = A_{nm}$ and thus it follows $\tilde{p}\left(\mathbf{x}_n|\boldsymbol{\xi}_n\right) \leq p\left(\mathbf{x}_n\right)$. Tipping (1999) and Bishop (2006) outline a variational EM algorithm that maximizes this lower bound and thus fit a latent trait analysis model.

However, the approximation of the log-likelihood obtained by using the variational approach is always less or equal than the true log-likelihood, so it may be advantageous to get a more accurate estimate the log-likelihood at the last step of the algorithm using Gauss-Hermite quadrature (Section 2.2.1); this is discussed further in Section 5.

6

2.2.4 Comparison of estimation methods

There are some advantages when using the variational approach to approximate the integral in the LTA likelihood. Firstly, the method involved avoids the need for user specified quantities like the number of quadrature points or the number of Monte Carlo samples. In addition, the variational approach involves iterating a series of closed-form updates that are easy to compute. Secondly, the algorithm usually converges considerably more quickly than in Gauss-Hermite quadrature and Monte Carlo sampling case, particularly so for large data sets and the approximation of the likelihood is more accurate as dimensionality increases, as the likelihood becomes more Gaussian in form. But, since the variational approach estimates the true likelihood function by using a lower bound, the approximation of the likelihood given by this method is always less or equal than the true value, so for low dimensional latent traits it may be better to use the Gauss-Hermite quadrature. Additionally, there are also the usual drawbacks of the EM algorithm: the results can vary because of the initialization of this algorithm and there is the risk of converging to a local maximum instead of the global maximum of the lower bound of the log-likelihood.

In Figure 1 we compare the response function estimated by the Gauss-Hermite quadrature with 6 quadrature points and the variational approach for a unidimensional latent trait model. The data consists of $M = 16$ binary variables and $N = 21574$ observations, and are fully described in Section 6.1. In this example the two methods bring to very similar results, as shown by the plot of the two response functions, and their absolute differences that are always < 0.1.

Furthermore, Tipping (1999) investigates the accuracy of the variational approach in a high dimensional context comparing the estimates given by the variational and the Monte Carlo sample approaches in terms of both computing time and errors.

3   Mixture of Latent Trait Analyzers

The mixture of latent trait analyzers (MLTA) model generalizes the latent class analysis and latent trait analysis by assuming that a set of N observations $\mathbf{x}_1, \ldots, \mathbf{x}_N$ comes from G different groups, and the behavior of the M categorical response variables given by each observation $\mathbf{x}_n$ ($n = 1, \ldots, N$) depends on both the group and the D dimensional continuous latent variable $\mathbf{y}_n$. Thus, the MLTA model is a mixture model for binary data but where observations are not necessarily conditionally independent given the group memberships. In fact, the observations within groups are modeled using a latent trait analysis model and thus dependence is accommodated.

Suppose that each observation comes from one of G groups and we have $\mathbf{z}_n = (z_{n1}, z_{n2}, \ldots, z_{nG})$
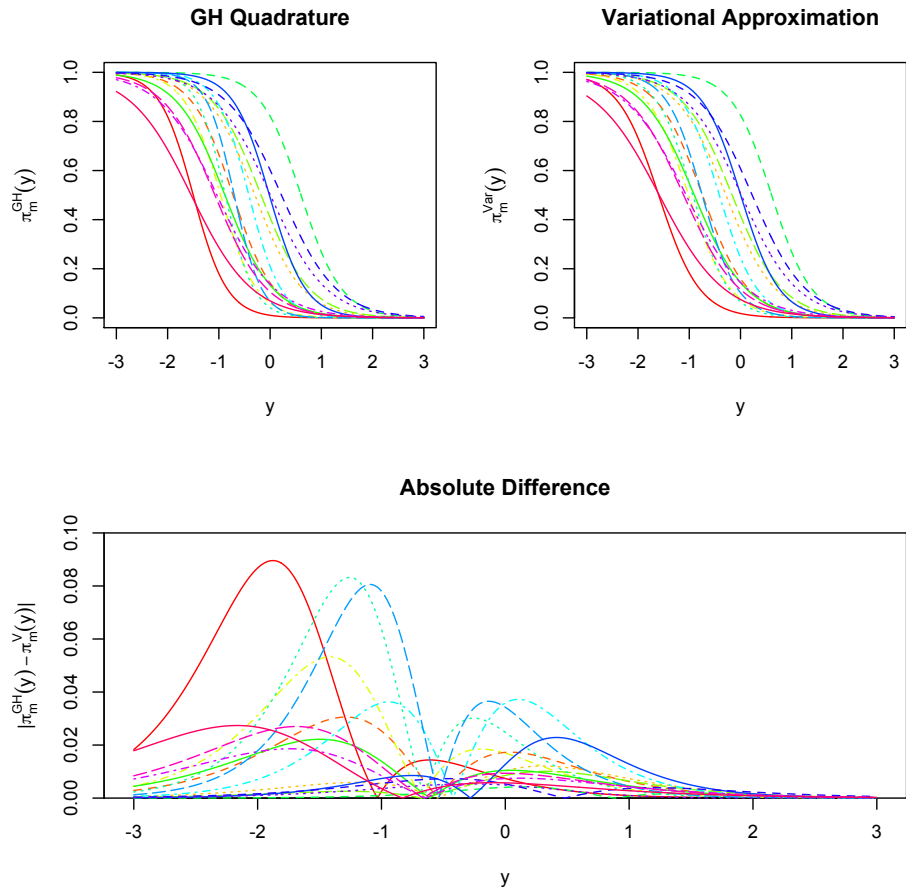
Figure 1: Response function for the Gauss-Hermite Quadrature and Variational Approach. The absolute difference of the response functions is also shown.

which is an indicator of the group membership. We assume that $\mathbf{z}_n \sim \text{Multinomial}(1, (\eta_1, \eta_2, \ldots, \eta_G))$ where η_g is the prior probability of a randomly chosen observation coming from the gth group ($\sum_{g'=1}^{G} \eta_{g'} = 1$ and $\eta_g \geq 0 \; \forall \, g = 1, \ldots, G$). Further, we assume that the conditional distribution of $\mathbf{x}_n$ given that the observation is from group g (ie. $z_{ng} = 1$) is a latent trait analysis model with parameters b_{mg} and $\mathbf{w}_{mg}$; this yields the mixture of latent trait analyzers (MLTA) model.

Thus, the MLTA model is of the form,

$$p(\mathbf{x}_n) = \sum_{g=1}^{G} \eta_g p\left(\mathbf{x}_n | z_{ng} = 1\right) = \sum_{g=1}^{G} \eta_g \int p\left(\mathbf{x}_n | \mathbf{y}_n, z_{ng} = 1\right) p(\mathbf{y}_n) \, d\mathbf{y}_n$$

where the conditional distribution given $\mathbf{y}_n$ and $z_{ng} = 1$ is

$$
\begin{aligned}
p\left(\mathbf{x}_n | \mathbf{y}_n, z_{ng} = 1\right) &= \prod_{m=1}^{M} p\left(x_{nm} | \mathbf{y}_n, z_{ng} = 1\right) \\
&= \prod_{m=1}^{M} \left(\pi_{mg}(\mathbf{y}_n)\right)^{x_{nm}} \left(1 - \pi_{mg}(\mathbf{y}_n)\right)^{1 - x_{nm}},
\end{aligned}
\tag{5}
$$

and the response function for each group is given by

$$\pi_{mg}(\mathbf{y}_n) = p\left(x_{nm} = 1 | \mathbf{y}_n, z_{ng} = 1\right) = \frac{1}{1 + \exp\left[-(b_{mg} + \mathbf{w}_{mg}^T \mathbf{y}_n)\right]}, \tag{6}$$

where b_{mg} and $\mathbf{w}_{mg}$ are the model parameters. In addition, it is assumed that the latent variable $\mathbf{y}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

The log-likelihood can be written as

$$\ell = \sum_{n=1}^{N} \log \left(\sum_{g=1}^{G} \eta_g \int \prod_{m=1}^{M} p\left(x_{nm} | \mathbf{y}_n, z_{ng} = 1\right) p(\mathbf{y}_n) \, d\mathbf{y}_n \right), \tag{7}$$

so, the model is a finite mixture model in which the component distributions are latent trait analysis models and the mixing proportions are $\eta_1, \eta_2, \ldots, \eta_G$.

3.1 Parsimonious Model

It is sometimes useful to use a more parsimonious response function than the one presented in (6); this is especially important when the data set is high dimensional, comes from several different groups and the continuous latent variable is high dimensional.

Similarly to the factor analysis setting (Bartholomew et al., 2011), $\frac{D \times (D-1)}{2}$ parameters are constrained because of the indeterminacy arising from all the possible rotations of $\mathbf{w}_g = (\mathbf{w}_{1g}, \ldots, \mathbf{w}_{Mg})$,

so the model in (6) involves $(G-1)+(G \times M)+G \times (M \times D - D \times (D-1)/2)$ free parameters, of which $G \times (M \times D - D \times (D-1)/2)$ are the parameters $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_G$.

If the $\mathbf{w}_g$ parameters are constrained to be the same in each group, a more parsimonious model would be:

$$\pi_{mg}(\mathbf{y}_n) = \frac{1}{1 + \exp\left[-(b_{mg} + \mathbf{w}_m^T \mathbf{y}_n)\right]}, \quad 0 \le \pi_{gm}(\mathbf{y}_n) \le 1, \tag{8}$$

that supposes that the $\mathbf{w}_{mg} \equiv \mathbf{w}_m$ for each group, so that it involves $(G-1)+(G \times M) + (M \times D - D \times (D-1)/2)$ free parameters.

3.2 Interpretation of Model Parameters

In the mixture of latent trait analyzers model η_g is the mixing proportion for the group g ($g = 1, \ldots, G$), that corresponds to the prior probability that a randomly chosen individual is in the g-th group. The behavior of the individuals within the group g is characterized by the parameters b_{mg} and $\mathbf{w}_{mg}$. In particular b_{mg} has a direct effect on the probability of a positive response to the variable m given by an individual in group g, through the relationship

$$\pi_{mg}(\mathbf{0}) = p(x_{nm} = 1 | \mathbf{y}_n = \mathbf{0}, z_{ng} = 1) = \frac{1}{1 + \exp(-b_{mg})}. \tag{9}$$

The value $\pi_{mg}(\mathbf{0})$ is the probability that the median individual in group g has a positive response for the variable m, since the continuous latent variable is distributed as a $\mathcal{N}(\mathbf{0}, \mathbf{I})$. A measure of the heterogeneity of the values of variable m within group g is given by the slope value $\mathbf{w}_{mg}$; the larger the value of $\mathbf{w}_{mg}$ the greater the differences in the probabilities of positive response in the variable m for observations from group g. The value of the slope parameters also account for the dependence between observed data variables. For example, if two variables m and m' yield a positive (negative) value for $\mathbf{w}_{mg}^T \mathbf{w}_{m'g}$ then they will both simultaneously have a probability of a positive outcome greater (lesser) than the median probability for group g more often than expected under local independence.

The quantity w_{dmg} can be used to calculate the correlation coefficient, within each group g, between the observed variables $\mathbf{x}_m$ and the latent variable $\mathbf{y}_d$ ($d = 1, \ldots, D$) which is given by its standardized value (Bartholomew et al., 2002),

$$w_{dmg}^* = \frac{w_{dmg}}{\sqrt{\sum_{d'=1}^{D} w_{d'mg}^2 + 1}}. \tag{10}$$

Another useful quantity for analyzing the dependence within groups is a version of *lift* (Brin et al., 1997). We use *lift* within each group to quantify the effect of the dependence on the

probability of two positive responses compared to the probability of two positive responses under an independence model. The *lift* is defined as,

$$lift = \frac{\mathbb{P}\left(x_{nm} = 1, x_{nk} = 1 | z_{ng} = 1\right)}{\mathbb{P}\left(x_{nm} = 1 | z_{ng} = 1\right)\mathbb{P}\left(x_{nk} = 1 | z_{ng} = 1\right)} \tag{11}$$

where $m = 1, 2, \ldots, M$ and $m \neq k$. Two independent positive responses have $lift = 1$: the more the variables are dependent, the further the value of the *lift* is from 1. It is possible to estimate it by using the Gauss-Hermite quadrature as,

$$lift \approx \frac{\sum_{q=1}^{Q} \pi_{mg}(\mathbf{y}_q)\pi_{kg}(\mathbf{y}_q)h(\mathbf{y}_q)}{\left(\sum_{q=1}^{Q} \pi_{mg}(\mathbf{y}_q)h(\mathbf{y}_q)\right)\left(\sum_{q=1}^{Q} \pi_{kg}(\mathbf{y}_q)h(\mathbf{y}_q)\right)}, \tag{12}$$

where $h(\mathbf{y}_1), \ldots, h(\mathbf{y}_Q)$ are the appropriate weights associated with the quadrature set of points $\mathbf{y}_1, \ldots, \mathbf{y}_Q$ and the parameters are estimated using the variational method (Section 3.4). Lift values that are much less than 1 are evidence of negative dependence within groups and lift values that are greater than 1 are evidence of positive dependence within groups.

Finally, the posterior distribution of the latent variable $\mathbf{y}_n$ conditional on the observation belonging to a particular group can be obtained from the model outputs (see Section 3.4) and the posterior mean estimates of these $\mathbf{y}_n$ scores can be used to interpret the latent variables within each group.

3.3 Related Models

The MLTA model has a lot of common characteristics with a number of models in the statistical and wider scientific literature.

The MLTA model can be seen as a discrete response variable analogue of the continuous response variable mixture of factor analyzers (MFA) model (Ghahramani and Hinton, 1997; McLachlan et al., 2003) and more recently developed parsimonious analogues of this model (McNicholas and Murphy, 2008, 2010; Baek et al., 2010). In these models, a discrete latent class variable accounts for grouping structure and a latent factor score accounts for correlation between response variables within groups. The component means in the MFA model are analogous to the intercepts in the MLTA model, the loading matrix is analogous to the slope parameters, the mixing proportions take an identical role in both models. Additionally, Muthén (2001) describes a general latent variable model for continuous outcome variables as used by the Mplus software and this model also uses both continuous and discrete latent variables in the model framework.

The Rasch model (Rasch, 1960) is a commonly used model for analyzing binary (and categorical data) and it has been widely used in educational statistics. The Rasch model has a very similar structure to the LTA model. In the Rasch model, the probability of a positive outcome is given as

$$\mathbb{P}(x_{nm} = 1|\beta_n, \delta_m) = \frac{\exp(\beta_n - \delta_m)}{1 + \exp(\beta_n - \delta_m)},$$

where the β_n values are called *ability* parameters and the δ_m values are called *difficulty* parameters and these correspond to the latent trait and the intercept parameters in the LTA model. Furthermore, the model makes a local independence assumption when constructing the probability

$$\mathbb{P}(\mathbf{x}_n|\beta_n, \boldsymbol{\delta}) = \prod_{m=1}^{M} \mathbb{P}(x_{nm}|\beta_n, \delta_m),$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)$. In some formulations of the Rasch model, the ability parameters are treated as fixed unknown parameters and in other formulations they are treated as random effects thus making the model equivalent to the univariate LTA model.

Consequently, the mixture of latent trait analyzers (MLTA) model has a similar structure to the mixture Rasch model (Rost, 1990; Rost and von Davier, 1995; von Davier and Yamamoto, 2007), where the response function is:

$$\mathbb{P}(x_{nm} = 1|\beta_n, \delta_{mg}, z_{ng} = 1) = \frac{\exp(\beta_n - \delta_{mg})}{1 + \exp(\beta_n - \delta_{mg})}.$$

However, as with the Rasch model, the mixture Rasch model usually includes a univariate ability parameter β_n that is the same across all the variables and groups, that would be equivalent to assuming that MLTA model $w_{1g} \equiv w_{2g} \equiv \ldots \equiv w_{Mg}$ and $w_{mg} \equiv w_m$ for all g.

In von Davier et al. (2007) they extend the mixture Rasch model in order to allow multivariate ability parameters:

$$\mathbb{P}(x_{nm} = 1|\mathbf{q}_m, \boldsymbol{\beta}_n, \delta_m, z_{ng} = 1) = \frac{\exp(\mathbf{q}_m^T \boldsymbol{\beta}_n - \delta_{mg})}{1 + \exp(\mathbf{q}_m^T \boldsymbol{\beta}_n - \delta_{mg})},$$

where $\mathbf{q}_m$ are variable-specific parameters, and $\boldsymbol{\beta}_n$ is the D-dimensional ability parameter. This is equivalent to a MLTA model with $w_{mg} \equiv w_m$ for all g, as in our parsimonious model.

Thus, many versions of the mixture Rasch model are included as special cases of the MLTA model. Again, the latent ability parameter is either assumed to be a fixed or random effect that varies across observations to accommodate different outcome probabilities; the latent ability parameter is the primary quantity of interest in many Rasch modeling applications.

Another model that has a very similar structure to the mixture Rasch model is the mixed latent trait (MLT) model described in Uebersax (1999), this model also has a univariate latent trait, uses a probit structure and numerical integration are required for computing the likelihood, so they can be difficult to apply to large heterogeneous data sets. Qu et al. (1996) and Hadgu and Qu (1998) use a two component model with a similar form to the MLT model in medical diagnosis.

Additionally, Uebersax (1999) also describes a probit latent class model (PLCA) which has a similar structure to the MLTA model, but this model uses a multivariate latent trait which has the same dimensionality as outcome variable. Vermunt and Magidson (2005) developed a latent class factor analysis (LCFA) model which uses discrete latent variables to develop a model with similar modeling aims to latent trait analysis, but which can be fitted in a much more computationally efficient manner.

Perhaps, the MLTA model proposed has closest connections to the family of models identified as multilevel mixture item response models (Vermunt, 2007). Key differences between our model and the multilevel mixture item response model are that we focus on a multivariate trait parameter (which is optional only in the model of Vermunt) and that we further also introduce a constrained parsimonious version of the model. In addition, we offer a computationally efficient alternative algorithm for fitting this model without the need to resort to numerical quadrature methods. The MLTA and multilevel mixture item response models can be estimated using the software Latent GOLD (Vermunt and Magidson, 2008), but the fact that this software uses quadrature methods for the numerical integration of continuous latent variables probably makes it less suitable for analyzing large datasets with underlying highdimensional latent trait structures.

3.4   Model Fitting

When fitting the MLTA model, the integral in the log-likelihood (7) cannot be solved analytically since it is exactly the same as in latent trait analysis (4). To obtain the approximation of the likelihood it is possible to use an EM algorithm:

1. *E-step:* Estimate z_{ng} as the approximate posterior probability that an individual with response vector $\mathbf{x}_n$ belongs to group g.

2. *M-step:* Estimate η_g as the proportion of observations in group g.

   Estimate the integral in the complete-data log-likelihood by using the variational approach.

   Estimate the parameters b_{mg} and $\mathbf{w}_{mg}$ for $m = 1, 2, \ldots, M$ and $g = 1, 2, \ldots, G$.

Also, estimate the approximate log-likelihood value.

3. Return to *step 1* until convergence is attained.

The advantages and the drawbacks of the different estimate methods for the likelihood are the same as in latent trait analysis (Section 2.2.4).

We further focus on the variational approach to approximate the log-likelihood because it can be efficiently implemented. Similarly to the latent trait analysis case, we introduce the variational parameters $\boldsymbol{\xi}_{ng} = (\xi_{n1g}, \ldots, \xi_{nMg})$ where $\xi_{nmg} \neq 0$ for all $m = 1, \ldots, M$ to approximate the logarithm of the component densities with a lower bound:

$$\mathcal{L}(\boldsymbol{\xi}_{ng}) = \log\left(\tilde{p}(\mathbf{x}_n | z_{ng} = 1, \boldsymbol{\xi}_{ng})\right)$$

$$= \log\left(\int \prod_{m=1}^{M} \tilde{p}\left(x_{nm} | \mathbf{y}_n, z_{ng} = 1, \xi_{nmg}\right) p(\mathbf{y}_n) \, d\mathbf{y}_n\right)$$

where the conditional distribution $p(x_{mn} |, \mathbf{y}_n, z_{ng} = 1)$ is approximated by,

$$\tilde{p}\left(x_{nm} | \mathbf{y}_n, z_{ng} = 1, \xi_{nmg}\right) = \sigma(\xi_{nmg}) \exp\left(\frac{A_{nmg} - \xi_{nmg}}{2} + \lambda(\xi_{nmg})\left(A_{nmg}^2 - \xi_{nmg}^2\right)\right),$$

where $\sigma(\xi_{nmg}) = (1 + \exp(-\xi_{nmg}))^{-1}$, $A_{nmg} = (2x_{nm} - 1)(b_{mg} + \mathbf{w}_{mg}^T \mathbf{y}_n)$ and $\lambda(\xi_{nmg}) = (\frac{1}{2} - \sigma(\xi_{nmg}))/2\xi_{nmg}$.

To obtain the approximation of the log-likelihood it is necessary to use a double EM algorithm on the $(i+1)$th iteration:

1. *E-step:* Estimate $z_{ng}^{(i+1)}$ with:

$$z_{ng}^{(i+1)} = \frac{\eta_g^{(i)} \exp\left(\mathcal{L}(\boldsymbol{\xi}_{ng}^{(i)})\right)}{\sum_{g'=1}^{G} \eta_{g'}^{(i)} \exp\left(\mathcal{L}(\boldsymbol{\xi}_{ng'}^{(i)})\right)}.$$

2. *M-step:* Estimate $\eta_g^{(i+1)}$ using:

$$\eta_g^{(i+1)} = \frac{\sum_{n=1}^{N} z_{ng}^{(i+1)}}{N}.$$

3. Estimate the likelihood:

   (a) *E-step:* Compute the latent posterior statistics for $\tilde{p}\left(\mathbf{y}_n | \mathbf{x}_n, z_{ng}^{(i+1)} = 1, \boldsymbol{\xi}_{ng}^{(i)}\right)$ which is a $\mathcal{N}(\boldsymbol{\mu}_{ng}^{(i+1)}, \mathbf{C}_{ng}^{(i+1)})$ density:

   $$\mathbf{C}_{ng}^{(i+1)} = \left[\mathbf{I} - 2\sum_{m=1}^{M} \lambda(\xi_{nmg}^{(i)})\mathbf{w}_{mg}^{(i)}\mathbf{w}_{mg}^{(i)T}\right]^{-1}$$

14

and

$$\boldsymbol{\mu}_{ng}^{(i+1)} = \mathbf{C}_{ng}^{(i+1)} \left[\sum_{m=1}^{M} \left(x_{nm} - \frac{1}{2} + 2\lambda(\xi_{nmg}^{(i)})b_{mg}^{(i)} \right) \mathbf{w}_{mg}^{(i)} \right].$$

(b) *M-step:* Optimize the variational parameter $\xi_{nmg}^{(i+1)}$ in order to make the approximation $\tilde{p}(\mathbf{x}_n | z_{ng}^{(i+1)} = 1, \boldsymbol{\xi}_{ng}^{(i+1)})$ as close as possible to $p(\mathbf{x}_n | z_{ng}^{(i+1)} = 1)$ for all $\mathbf{x}_n$ using:

$$\xi_{nmg}^{(i+1)2} = \mathbf{w}_{mg}^{(i)T} \left(\mathbf{C}_{ng}^{(i+1)} + \boldsymbol{\mu}_{ng}^{(i+1)} \boldsymbol{\mu}_{ng}^{(i+1)T} \right) \mathbf{w}_{mg}^{(i)} + 2b_{mg}^{(i)} \mathbf{w}_{mg}^{(i)T} \boldsymbol{\mu}_{ng}^{(i+1)} + b_{mg}^{(i)2},$$

since $\tilde{p}(x_{nm} | \mathbf{y}_n, z_{ng} = 1, \xi_{nmg})$ is symmetric in ξ_{nmg} choosing either the positive or the negative root of ξ_{nmg}^2 yields to the same results.

(c) Optimise the model parameters $\mathbf{w}_{mg}^{(i+1)}$ and $b_{mg}^{(i+1)}$ in order to increase the approximate likelihood $\tilde{p}\left(\mathbf{x}_1, \ldots, \mathbf{x}_N | \boldsymbol{\xi}_{11}^{(i+1)}, \ldots, \boldsymbol{\xi}_{NG}^{(i+1)}\right)$ using:

$$\hat{\mathbf{w}}_{mg}^{(i+1)} = - \left[2 \sum_{n=1}^{N} z_{ng}^{(i+1)} \lambda(\xi_{nmg}^{(i+1)}) \mathbb{E}[\hat{\mathbf{y}}_n \hat{\mathbf{y}}_n^T]_g^{(i+1)} \right]^{-1} \left[\sum_{n=1}^{N} z_{ng}^{(i+1)} \left(x_{nm} - \frac{1}{2} \right) \hat{\boldsymbol{\mu}}_{ng}^{(i+1)} \right],$$

where $\hat{\mathbf{w}}_{mg}^{(i+1)} = (\mathbf{w}_{mg}^{(i+1)T}, b_{mg}^{(i+1)})^T$, $\hat{\boldsymbol{\mu}}_{ng}^{(i+1)} = (\boldsymbol{\mu}_{ng}^{(i+1)T}, 1)^T$, and

$$\mathbb{E}[\hat{\mathbf{y}}_n \hat{\mathbf{y}}_n^T]_g^{(i+1)} = \begin{bmatrix} \mathbf{C}_{ng}^{(i+1)} + \boldsymbol{\mu}_{ng}^{(i+1)} \boldsymbol{\mu}_{ng}^{(i+1)T} & \boldsymbol{\mu}_{ng}^{(i+1)} \\ \boldsymbol{\mu}_{ng}^{(i+1)T} & 1 \end{bmatrix}.$$

(d) Estimate the lower bound:

$$\mathcal{L}(\boldsymbol{\xi}_{ng}^{(i+1)}) = \sum_{m=1}^{M} \left[\log\left(\sigma(\xi_{nmg}^{(i+1)})\right) - \frac{\xi_{nmg}^{(i+1)}}{2} - \lambda(\xi_{nmg}^{(i+1)})\xi_{nmg}^{(i+1)2} + \left(x_{nm} - \frac{1}{2} \right) b_{mg}^{(i+1)} + \lambda(\xi_{nmg}^{(i+1)}) b_{mg}^{(i+1)2} \right]$$
$$+ \frac{\log |\mathbf{C}_{ng}^{(i+1)}|}{2} + \frac{\boldsymbol{\mu}_{ng}^{(i+1)T} [\mathbf{C}_{ng}^{(i+1)}]^{-1} \boldsymbol{\mu}_{ng}^{(i+1)}}{2},$$

and the log-likelihood:

$$\ell^{(i)} \approx \sum_{n=1}^{N} \log \left(\sum_{g=1}^{G} \eta_g^{(i+1)} \exp(\mathcal{L}(\boldsymbol{\xi}_{ng}^{(i+1)})) \right).$$

4. Return to *step* 1 until convergence is attained.

5. Estimate the log-likelihood by using the Gauss-Hermite quadrature:

$$\ell \approx \sum_{n=1}^{N} \log \left(\sum_{g=1}^{G} \eta_g \sum_{q=1}^{Q} p(\mathbf{x}_n | \mathbf{y}_q, z_{ng} = 1) h(\mathbf{y}_q) \right), \tag{13}$$

where $h(\mathbf{y}_1), \ldots, h(\mathbf{y}_Q)$ are the appropriate weights associated to the quadrature set of points $\mathbf{y}_1, \ldots, \mathbf{y}_Q$ and $p(\mathbf{x}_n | \mathbf{y}_q, z_{ng} = 1)$ is given by (5) and (6).

To fit the parsimonious model outlined in Section 3.1 it is possible to use the variational approach using the EM algorithm described above, except for the estimate of the model parameters at *step* 3c:

$$\hat{\mathbf{w}}_m^{(i+1)} = - \left[2\mathbf{K}_m^{(i+1)}\right]^{-1} \boldsymbol{\gamma}_m^{(i+1)},$$

where

$$\hat{\mathbf{w}}_m^{(i+1)} = (\mathbf{w}_m^{(i+1)T}, b_{m1}^{(i+1)}, \ldots, b_{mG}^{(i+1)})^T, \quad \boldsymbol{\gamma}_m^{(i+1)} = \begin{bmatrix} \sum_{g=1}^{G} \sum_{n=1}^{N} z_{ng}^{(i+1)} \left(x_{nm} - \frac{1}{2}\right) \boldsymbol{\mu}_{ng}^{(i+1)} \\ \sum_{n=1}^{N} z_{n1}^{(i+1)} \left(x_{nm} - \frac{1}{2}\right) \\ \vdots \\ \sum_{n=1}^{N} z_{nG}^{(i+1)} \left(x_{nm} - \frac{1}{2}\right) \end{bmatrix}$$

and,

$$\mathbf{K}_m^{(i+1)} = \begin{bmatrix} \sum_{g=1}^{G} \sum_{n=1}^{N} z_{ng}^{(i+1)} \lambda(\xi_{nmg}^{(i+1)}) \mathbb{E}\left[\mathbf{y}_n \mathbf{y}_n^T\right]_g^{(i+1)} & \sum_{n=1}^{N} z_{n1}^{(i+1)} \lambda(\xi_{nm1}^{(i+1)}) \boldsymbol{\mu}_{n1}^{(i+1)} & \cdots & \sum_{n=1}^{N} z_{nG}^{(i+1)} \lambda(\xi_{nmG}^{(i+1)}) \boldsymbol{\mu}_{nG}^{(i+1)} \\ \sum_{n=1}^{N} z_{n1}^{(i+1)} \lambda(\xi_{nm1}^{(i+1)}) \boldsymbol{\mu}_{n1}^{(i+1)T} & \sum_{n=1}^{N} z_{n1}^{(i+1)} \lambda(\xi_{nm1}^{(i+1)}) & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ \sum_{n=1}^{N} z_{nG}^{(i+1)} \lambda(\xi_{nmG}^{(i+1)}) \boldsymbol{\mu}_{nG}^{(i+1)T} & 0 & 0 & \sum_{n=1}^{N} z_{nG}^{(i+1)} \lambda(\xi_{nmG}^{(i+1)}) \end{bmatrix}$$

The derivation of these parameter estimates is given in Section 1 of the supplementary material.

3.5   Model Identifiability

Model identifiability is an important issue for a model involving many parameters. Goodman (1974), Bartholomew et al. (2011) and Dean and Raftery (2010) give a detailed explanation of model identifiability in latent class analysis and Bartholomew (1980) introduces this issue in the latent trait analysis context. In Allman et al. (2009) they argue that the classical definition of identifiability is too strong for a lot of latent variable models, and they introduce the concept of "generic identifiability" that implies that the set of points for which identifiability does not hold has measure zero. They explore "generic identifiability" for different models, including LCA.

A necessary condition for model identifiability is that the number of the free estimated parameters not exceed the number of possible data patterns. Nevertheless this condition is not sufficient as the actual information in a dataset can be less depending of the size or the frequency of pattern occurrences within the dataset.

16

As with the loading matrices in mixture models with a factor analytic structure (eg. Ghahramani and Hinton, 1997; McNicholas and Murphy, 2008) the slope parameters $\mathbf{w}_g$ are only identifiable up to a rotation of the factors. The rotational freedom of the factor scores is important when determining the number of parameters in the model (Section 3.1 and Section 4).

In addition, model identifiability holds if the observed information matrix is full rank. As a result possible non-identifiability can manifest itself through high standard errors of the parameter estimates. Another empirical method that can be used to assess non-identifiability consists of checking whether the same maximized likelihood value is reached with different estimates of the model parameter values when starting the EM algorithm from different values.

We found that these checks for identifiability were all satisfied in the empirical examples discussed in Section 6.

4   Model Selection

The Bayesian Information Criterion (BIC) (Schwarz (1978)) can be used to select the model,

$$\text{BIC} = -2\ell + k \log(N),$$

where k is the number of free parameters in the model and N is the number of observations. The model with the lower value of BIC is preferable. It is important to remember that its value could be overestimated if the log-likelihood is approximated by using the variational approach; hence, the proposal to use Gauss Hermite quadrature to evaluate the maximized log-likelihood for model selection purposes.

In the context of MLTA, the values of G, D and whether the $\mathbf{w}_g$ values are constrained to be equal across groups need to be determined.

In the mixture of latent trait analyzers context (and more widely within finite mixture models) the BIC penalizes too much the models with high D and/or high G. The BIC as defined by Schwarz (1978) implicitly assumes that all the parameter estimates depend on the entire set of N observations, but in MLTA the estimates of $\mathbf{b}_g$ and $\mathbf{w}_g$ depend just on the observations that belong to group g. So, an alternative penalized BIC is given as

$$\text{BIC}^* = -2\ell + (G-1)\log(N) + k^* \sum_{g=1}^{G} \log(\eta_g N)$$

$$= \text{BIC} + k^* \sum_{g=1}^{G} \log(\eta_g),$$

where k^* is the number of free parameters depending on each group of observations. This penalty penalizes parameters to a lesser extent than BIC because only the estimated number of observations involved in the estimation of each parameter is used in the penalty. This version of BIC has previously been proposed by Steele (2002) and similar criteria have been proposed by Pauler (1998) and Raftery et al. (2007). It is worth noting that BIC* will behave in a similar way to BIC for large sample sizes. Hence, it can be seen as a small sample adjustment to BIC.

The Pearson's χ^2-test can be used to check the goodness of the model fit. The χ^2 statistic is calculated as

$$
\chi^2 = \sum_{p=1}^{P} \frac{(\mathrm{Observed}_p - \mathrm{Expected}_p)^2}{\mathrm{Expected}_p}
$$
$$
= \sum_{\underline{n}=1}^{\underline{N}} \frac{(\mathrm{Observed}_{\underline{n}} - \mathrm{Expected}_{\underline{n}})^2}{\mathrm{Expected}_{\underline{n}}} + \left(N - \sum_{\underline{n}=1}^{\underline{N}} \mathrm{Expected}_{\underline{n}} \right),
$$

where $P = 2^M$ is the total number of possible patterns, $\underline{N}$ is the number of observed patterns and $\mathrm{Observed}_{\underline{n}}$ and $\mathrm{Expected}_{\underline{n}} = Np(\mathbf{x}_{\underline{n}})$ represent the observed and expected frequencies for the $\underline{n}$-th response pattern, respectively. Under the null hypothesis the χ^2 statistic is asymptotically distributed as a χ^2_{P-k-1} as $N \to \infty$. If M is large it is common to have a large number of very small counts and the Pearson's χ^2-test is not applicable. In this case, Erosheva et al. (2007) suggest the truncated SSPR criterion to examine deviations between expected and observed frequencies via the sum of squared Pearson residual only for the patterns with large observed frequencies.

5 Computational Aspects

The estimates of the parameters in the latent class analysis models ($D = 0$) are the exact maximum likelihood estimates. Since the dimensionality of the data in the two data sets is large, the parameters for the model with $D \geq 1$ are estimated by using the variational approach and the log-likelihood has been calculated at the last step of the algorithm by using the Gauss-Hermite quadrature, with 5 quadrature points per dimension; the results were not sensitive to the number of quadrature points but the computational time was very heavily dependent on this number.

The categorical latent variables $\mathbf{z}_n$ ($n = 1, \ldots, N$) have been initialized by randomly assigning each observation to one of the G possible groups. The variational parameters ξ_{nmg} ($n = 1, \ldots, N$, $m = 1, \ldots, M$, $g = 1, \ldots, G$) are initialized to be equal to 20, this implies the initial approximation of the conditional distribution to be very close to 0 and it reduces the depen-

dence of the final estimates on the initializing values. The model parameters b_{mg} and $\mathbf{w}_{mg}$ have been initialized by random generated numbers from a $\mathcal{N}(0,1)$. Ten random starts of the algorithm were used and the solution with the maximum likelihood value was selected.

The standard errors of the model parameter have been calculated using the jackknife method (Efron, 1981). It is worth noting that when employing this method the estimates of the parameter without the n-th observation ($n = 1, \ldots, N$) can be obtained in just a few iterations.

Since the EM algorithm is linearly convergent, a criterion based on the Aitken acceleration (McLachlan and Peel, 2000) has been used to determine the convergence of the algorithm. The EM algorithm has been stopped when

$$\|\ell_A^{(i+1)} - \ell_A^{(i)}\| < \text{tol}$$

where i is the iteration, $\text{tol} = 10^{-2}$ is the desired tolerance,

$$\ell_A^{(i+1)} = \ell^{(i)} + \frac{1}{1 - a^{(i)}} \left(\ell^{(i+1)} - \ell^{(i)} \right) \qquad \text{and} \qquad a^{(i)} = \frac{\ell^{(i+1)} - \ell^{(i)}}{\ell^{(i)} - \ell^{(i-1)}}.$$

6 Applications

6.1 National Long Term Care Survey (NLTCS)

Erosheva (2002, 2003, 2004) and Erosheva et al. (2007) report on 16 binary outcome variables recorded for 21574 elderly (age 65 and above) people who took part in the National Long-Term Care Survey in the years 1982, 1984, 1989 and 1994. The outcome variables record the level of disability and can be divided in two subsets: the "activities of daily living" (ADLs) and "instrumental activities of daily living" (IADLs). The first subset is composed of the first six variables which include basic activities of hygiene and personal care: *(1)* eating, *(2)* getting in/out of bed, *(3)* getting around inside, *(4)* dressing, *(5)* bathing, *(6)* using the toilet. The second subset concern the basic activities necessary to reside in the community: *(7)* doing heavy house work, *(8)* doing light house work, *(9)* doing laundry, *(10)* cooking, *(11)* grocery shopping, *(12)* getting about outside, *(13)* travelling, *(14)* managing money, *(15)* taking medicine, *(16)* telephoning. The responses are coded as $1 =$ disabled and $0 =$ able.

The MLTA model is fitted to these data for $D = 0, 1, 2, 3$ and $G = 1, 2, \ldots, 11$; the $D = 0$ case corresponds to the LCA model and the $G = 1$ & $D > 0$ case corresponds to the LTA model.

Table 1 records the log-likelihood evaluated for the parameter estimates found using the algorithm outlined in Section 2.2.3. Estimates of the log-likelihood and the lower bound are reported

to emphasize the importance of estimating the log-likelihood at the last step of the variational approach by using the Gauss-Hermite quadrature instead of the lower bound values.

Table 1: Approximated log-likelihood and lower bound (in parentheses).

	$D = 0$	$D = 1$		$D = 2$		$D = 3$	
$G = 1$	-200085.10	-140318.06	(-145827.00)	-136169.53	(-144483.30)	-136075.66	(-144483.30)
$G = 2$	-152527.30	-135301.29	(-139930.20)	-134273.79	(-139283.00)	-134275.46	(-139091.30)
$G = 3$	-141277.10	-134362.61	(-136349.10)	-133025.27	(-136080.80)	-133008.17	(-136066.50)
$G = 4$	-137464.20	-133120.36	(-134540.70)	-131839.77	(-134253.60)	-132116.82	(-134145.00)
$G = 5$	-135216.20	-131813.29	(-133203.40)	-131505.23	(-133061.10)	-131393.42	(-132919.80)
$G = 6$	-133643.80	-131396.59	(-132261.50)	-131154.94	(-132161.60)	-130992.52	(-132123.00)
$G = 7$	-132659.70	-131120.79	(-131840.30)	-130729.39	(-131785.80)	-130607.37	(-131731.10)
$G = 8$	-132202.90	-130708.20	(-131565.90)	-130450.55	(-131146.60)	-130403.20	(-131106.60)
$G = 9$	-131367.70	-130342.81	(-130866.20)	-130164.32	(-130855.20)	-130155.19	(-130798.00)
$G = 10$	-131155.90	-130135.91	(-130806.60)	-130049.64	(-130681.20)	-129936.33	(-130544.50)
$G = 11$	-130922.60	-130110.22	(-130574.40)	-129860.74	(-130475.10)	-129881.83	(-130404.50)

The estimated BIC and BIC* have been used to select the best model (Table 2), both criteria indicate that the model with 10 groups and a one-dimensional latent variable as the best model. Fienberg et al. (2009) show that the LCA model that minimizes the BIC has nineteen latent classes, so the mixture of latent trait analyzers suggests that there are much fewer groups. The nineteen latent class model has a lower BIC (BIC=262165.07) but they argue that this number of classes is not sensible in the context of the application. This suggests that the LCA is using multiple groups to account for dependence and that the latent classes do not necessarily correspond to data clusters.

Since there is a large number of response patterns with a very small number of observations (of all the $2^{16} = 65536$ possible response patterns, 62384 (95.2%) contain zero counts and only 481 (0.7%) contain more than 5 counts), the Pearson's χ^2 test is not applicable and the truncated SSPR criterion has been used to test the goodness of fit of the model. Three different levels of truncation are shown in Table 3.

The best model selected by the BIC and the BIC* is one of the best fits as indicated by the SSPR over all the three levels of truncation. Table 1 of the supplementary material shows a comparison of the observed and the expected frequencies for the response patterns with more than 100 observations under the best model selected. The table shows that there is a close match between the observed and expected frequencies under this model.

Table 2: The estimated BIC (left) and BIC* (right) for the models with $D = 0, 1, 2, 3$ and $G = 1, 2, \ldots, 11$.

	$D = 0$	$D = 1$	$D = 2$	$D = 3$		$D = 0$	$D = 1$	$D = 2$	$D = 3$
$G = 1$	400329.84	280955.46	272808.09	272760.05	$G = 1$	400329.84	280955.46	272808.09	272760.05
$G = 2$	305383.99	271251.24	269495.61	269778.37	$G = 2$	305360.21	271204.68	269427.88	269690.47
$G = 3$	283053.38	269703.18	267477.56	267862.51	$G = 3$	282997.58	269596.55	267322.64	267661.41
$G = 4$	275597.01	267548.00	265585.58	266698.51	$G = 4$	275504.40	267367.43	265322.50	266345.65
$G = 5$	271270.65	265263.18	265395.51	265870.43	$G = 5$	271136.22	264993.58	265006.92	265366.66
$G = 6$	268295.46	264759.09	265173.93	265687.34	$G = 6$	268113.98	264403.40	264614.73	264973.09
$G = 7$	266496.97	264536.80	264801.82	265535.74	$G = 7$	266267.39	264069.53	264102.10	264643.31
$G = 8$	264882.64	264040.94	264723.15	265746.12	$G = 8$	264591.64	263464.76	263877.58	264649.68
$G = 9$	264252.28	263639.47	264629.69	265868.82	$G = 9$	263910.91	262952.98	263618.30	264544.87
$G = 10$	263998.39	**263554.99**	264879.33	266049.81	$G = 10$	263600.50	**262766.36**	263684.61	264524.20
$G = 11$	263701.27	263832.93	264980.55	266559.53	$G = 11$	263242.58	262921.59	263581.68	264754.70

From Table 2 and 3 it is also possible to see how the mixture of latent trait analyzers model is more appropriate than the latent class analysis and the latent trait analysis, in terms both of BIC and goodness of fit.

Table 3: The sum of squared of Pearson residuals for different levels of truncation (≥ 100, ≥ 25, ≥ 10).

	observed frequencies ≥ 100				observed frequencies ≥ 25				observed frequencies ≥ 10			
	$D = 0$	$D = 1$	$D = 2$	$D = 3$	$D = 0$	$D = 1$	$D = 2$	$D = 3$	$D = 0$	$D = 1$	$D = 2$	$D = 3$
$G = 1$	6.1e+09	2441	1924	4082	6.3e+09	20524	21513	25795	6.3e+09	38927	25130	30009
$G = 2$	80199	1600	1318	1718	129642	4614	3590	4822	182265	8126	6614	9654
$G = 3$	5304	1470	1410	1597	16588	4610	2974	3318	61331	7610	4898	5814
$G = 4$	4875	1439	638	806	10762	3060	1874	2530	17405	5498	2993	4609
$G = 5$	2717	598	490	545	6321	1788	1498	1683	10886	3414	3037	3255
$G = 6$	1434	323	533	462	3901	1423	1787	1530	6652	2498	2758	2817
$G = 7$	561	561	561	418	2412	2412	2412	1450	4808	4808	4808	2468
$G = 8$	413	413	413	234	1837	1837	1837	1228	3792	3792	3792	2115
$G = 9$	417	253	183	229	1647	987	938	1054	3010	1725	1534	1761
$G = 10$	348	**160**	292	**150**	1348	**723**	827	946	2695	**1367**	1495	1558
$G = 11$	280	196	171	**118**	1336	872	760	**700**	2246	1518	**1348**	1481

The parameter estimates and standard errors for the selected model are reported in Table 4. The standardized values w_{mg}^{*} and median probabilities $\pi_{mg}(0)$ are also reported to aid interpretation of the groups found by this model.

Table 4: Parameter estimates (*standard errors*) for the selected model ($G = 10$, $D = 1$)

$g = 1$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	-3.16 (0.11)	0.02 (0.04)	0.02	0.04
$m = 2$	-0.97 (0.08)	1.11 (0.06)	0.74	0.31
$m = 3$	0.38 (0.09)	1.36 (0.05)	0.81	0.57
$m = 4$	-1.46 (0.09)	0.38 (0.07)	0.36	0.19
$m = 5$	0.64 (0.08)	0.64 (0.07)	0.54	0.65
$m = 6$	-1.19 (0.08)	0.73 (0.07)	0.59	0.25
$m = 7$	5.95 (0.08)	-0.05 (0.02)	-0.05	1
$m = 8$	0.69 (0.08)	-0.14 (0.07)	-0.14	0.67
$m = 9$	2.84 (0.11)	-0.12 (0.05)	-0.12	0.94
$m = 10$	1.54 (0.10)	0.31 (0.07)	0.30	0.82
$m = 11$	3.10 (0.12)	0.29 (0.07)	0.28	0.96
$m = 12$	1.91 (0.10)	1.00 (0.08)	0.71	0.84
$m = 13$	1.62 (0.09)	0.66 (0.08)	0.55	0.82
$m = 14$	0.11 (0.09)	1.14 (0.06)	0.75	0.52
$m = 15$	-0.38 (0.08)	0.89 (0.06)	0.67	0.42
$m = 16$	-0.98 (0.08)	0.93 (0.06)	0.68	0.30
	$\eta_g = 0.070$ (0.002)			

$g = 2$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	-2.89 (0.12)	0.96 (0.12)	0.69	0.07
$m = 2$	-4.53 (0.12)	0.05 (0.06)	0.05	0.01
$m = 3$	-2.81 (0.15)	-0.01 (0.07)	-0.01	0.06
$m = 4$	-1.90 (0.12)	1.50 (0.11)	0.83	0.20
$m = 5$	-0.49 (0.10)	1.42 (0.09)	0.82	0.41
$m = 6$	-2.74 (0.12)	1.05 (0.12)	0.72	0.09
$m = 7$	1.78 (0.11)	1.70 (0.09)	0.86	0.77
$m = 8$	-1.54 (0.11)	1.77 (0.08)	0.87	0.26
$m = 9$	0.61 (0.11)	1.80 (0.08)	0.87	0.60
$m = 10$	0.69 (0.11)	1.73 (0.08)	0.87	0.61
$m = 11$	2.30 (0.13)	0.82 (0.12)	0.63	0.89
$m = 12$	-0.28 (0.09)	0.67 (0.09)	0.56	0.44
$m = 13$	1.72 (0.11)	0.08 (0.10)	0.08	0.85
$m = 14$	2.02 (0.14)	0.25 (0.07)	0.25	0.88
$m = 15$	0.52 (0.09)	0.83 (0.09)	0.64	0.61
$m = 16$	0.31 (0.09)	0.47 (0.09)	0.42	0.57
	$\eta_g = 0.048$ (0.002)			

$g = 3$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	-5.68 (0.06)	0.00 (0.00)	0.00	0
$m = 2$	-0.26 (0.05)	0.13 (0.04)	0.13	0.44
$m = 3$	2.37 (0.07)	-0.08 (0.02)	-0.08	0.91
$m = 4$	-2.72 (0.08)	0.07 (0.02)	0.07	0.06
$m = 5$	0.38 (0.05)	0.83 (0.02)	0.64	0.59
$m = 6$	-0.92 (0.06)	0.67 (0.04)	0.56	0.29
$m = 7$	2.51 (0.08)	0.11 (0.03)	0.11	0.92
$m = 8$	-3.44 (0.07)	-0.01 (0.01)	-0.01	0.03
$m = 9$	-0.37 (0.05)	0.23 (0.05)	0.22	0.41
$m = 10$	-2.21 (0.07)	0.01 (0.02)	0.01	0.10
$m = 11$	1.68 (0.07)	0.05 (0.03)	0.05	0.84
$m = 12$	4.39 (0.06)	0.00 (0.01)	0.00	0.99
$m = 13$	1.60 (0.07)	0.00 (0.03)	0.00	0.83
$m = 14$	-1.59 (0.06)	-0.03 (0.03)	-0.03	0.17
$m = 15$	-2.29 (0.07)	0.01 (0.02)	0.01	0.09
$m = 16$	-3.09 (0.08)	-0.07 (0.02)	-0.07	0.04
	$\eta_g = 0.108$ (0.003)			

$g = 4$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	0.27 (0.13)	-0.84 (0.13)	-0.64	0.56
$m = 2$	2.51 (0.25)	-0.82 (0.16)	-0.63	0.91
$m = 3$	1.76 (0.20)	-0.23 (0.18)	-0.22	0.85
$m = 4$	2.32 (0.21)	-0.87 (0.19)	-0.66	0.89
$m = 5$	4.01 (0.23)	-0.29 (0.13)	-0.27	0.98
$m = 6$	1.84 (0.20)	-0.30 (0.17)	-0.29	0.86
$m = 7$	1.81 (0.16)	-1.73 (0.17)	-0.87	0.77
$m = 8$	-0.95 (0.17)	-2.37 (0.11)	-0.92	0.38
$m = 9$	-0.18 (0.17)	-2.63 (0.10)	-0.93	0.47
$m = 10$	-0.25 (0.16)	-2.47 (0.10)	-0.93	0.46
$m = 11$	0.22 (0.16)	-1.79 (0.15)	-0.87	0.53
$m = 12$	-1.00 (0.22)	0.69 (0.14)	0.57	0.29
$m = 13$	-0.80 (0.17)	0.13 (0.14)	0.13	0.31
$m = 14$	-0.20 (0.17)	-1.27 (0.14)	-0.79	0.46
$m = 15$	0.32 (0.14)	-0.90 (0.13)	-0.67	0.57
$m = 16$	-0.66 (0.17)	-1.59 (0.14)	-0.85	0.39
	$\eta_g = 0.022$ (0.002)			

$g = 5$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	-4.91 (0.05)	0.02 (0.01)	0.02	0.01
$m = 2$	-4.07 (0.05)	0.02 (0.01)	0.02	0.02
$m = 3$	-4.61 (0.02)	0.00 (0.01)	0.00	0.01
$m = 4$	-3.23 (0.06)	-0.01 (0.02)	-0.01	0.04
$m = 5$	-0.77 (0.05)	-0.36 (0.04)	-0.34	0.32
$m = 6$	-2.78 (0.07)	-0.07 (0.03)	-0.07	0.06
$m = 7$	1.85 (0.05)	0.22 (0.03)	0.21	0.86
$m = 8$	-2.92 (0.07)	0.22 (0.03)	0.22	0.05
$m = 9$	-0.83 (0.05)	0.41 (0.04)	0.38	0.31
$m = 10$	-2.42 (0.06)	0.13 (0.03)	0.13	0.08
$m = 11$	0.67 (0.05)	-0.72 (0.03)	-0.59	0.65
$m = 12$	-0.24 (0.05)	-0.79 (0.03)	-0.62	0.45
$m = 13$	0.82 (0.05)	-0.95 (0.03)	-0.69	0.68
$m = 14$	-1.66 (0.06)	-0.25 (0.04)	-0.24	0.16
$m = 15$	-2.25 (0.06)	0.06 (0.03)	0.06	0.10
$m = 16$	-3.08 (0.07)	0.06 (0.02)	0.06	0.04
	$\eta_g = 0.149$ (0.003)			

$g = 6$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	0.79 (0.05)	-0.69 (0.04)	-0.57	0.68
$m = 2$	3.27 (0.08)	-0.05 (0.02)	-0.05	0.96
$m = 3$	3.88 (0.08)	0.03 (0.02)	0.03	0.98
$m = 4$	3.34 (0.06)	-0.08 (0.02)	-0.08	0.97
$m = 5$	4.90 (0.07)	0.00 (0.01)	0.00	0.99
$m = 6$	2.11 (0.07)	0.09 (0.04)	0.09	0.89
$m = 7$	6.41 (0.06)	0.01 (0.01)	0.01	1
$m = 8$	2.90 (0.07)	-0.02 (0.02)	-0.02	0.95
$m = 9$	4.60 (0.07)	0.00 (0.01)	0.00	0.99
$m = 10$	4.25 (0.07)	-0.03 (0.01)	-0.03	0.99
$m = 11$	4.48 (0.08)	-0.03 (0.02)	-0.03	0.99
$m = 12$	4.34 (0.07)	0.03 (0.02)	0.03	0.99
$m = 13$	2.52 (0.07)	-0.03 (0.03)	-0.03	0.93
$m = 14$	1.32 (0.06)	-1.04 (0.04)	-0.72	0.76
$m = 15$	1.56 (0.06)	-0.68 (0.05)	-0.56	0.81
$m = 16$	0.47 (0.06)	-1.23 (0.03)	-0.78	0.60
	$\eta_g = 0.112$ (0.002)			

$g = 7$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	-7.61 (0.19)	0.00 (0.00)	0.00	0
$m = 2$	-8.79 (0.26)	0.00 (0.00)	0.00	0
$m = 3$	-10.13 (0.00)	0.00 (0.00)	0.00	0
$m = 4$	-6.99 (0.10)	0.00 (0.00)	0.00	0
$m = 5$	-6.50 (0.26)	0.00 (0.00)	0.00	0
$m = 6$	-8.72 (0.33)	0.00 (0.00)	0.00	0
$m = 7$	-7.35 (0.19)	0.00 (0.00)	0.00	0
$m = 8$	-11.07 (0.10)	0.00 (0.00)	0.00	0
$m = 9$	-8.88 (0.04)	0.00 (0.00)	0.00	0
$m = 10$	-8.16 (0.04)	0.00 (0.00)	0.00	0
$m = 11$	-9.60 (0.04)	0.00 (0.00)	0.00	0
$m = 12$	-3.91 (0.07)	0.00 (0.00)	0.00	0.02
$m = 13$	-4.28 (0.03)	0.00 (0.00)	0.00	0.01
$m = 14$	-6.11 (0.09)	0.00 (0.00)	0.00	0
$m = 15$	-7.95 (0.14)	0.00 (0.00)	0.00	0
$m = 16$	-4.03 (0.04)	0.00 (0.00)	0.00	0.02
	$\eta_g = 0.138$ (0.003)			

$g = 8$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	-5.87 (0.05)	0.00 (0.00)	0.00	0
$m = 2$	-4.43 (0.03)	0.00 (0.00)	0.00	0.01
$m = 3$	-4.81 (0.02)	0.00 (0.00)	0.00	0.01
$m = 4$	-3.55 (0.05)	0.00 (0.00)	0.00	0.03
$m = 5$	-1.55 (0.04)	0.00 (0.00)	0.00	0.17
$m = 6$	-3.29 (0.06)	0.00 (0.00)	0.00	0.04
$m = 7$	-0.07 (0.03)	0.00 (0.00)	0.00	0.48
$m = 8$	-8.32 (0.25)	0.00 (0.00)	0.00	0
$m = 9$	-5.96 (0.03)	0.00 (0.00)	0.00	0
$m = 10$	-5.48 (0.05)	0.00 (0.00)	0.00	0
$m = 11$	-3.23 (0.05)	0.00 (0.00)	0.00	0.04
$m = 12$	-1.38 (0.04)	0.00 (0.00)	0.00	0.20
$m = 13$	-1.94 (0.05)	0.00 (0.00)	0.00	0.13
$m = 14$	-3.70 (0.06)	0.00 (0.00)	0.00	0.02
$m = 15$	-2.92 (0.05)	0.00 (0.00)	0.00	0.05
$m = 16$	-3.55 (0.06)	0.00 (0.00)	0.00	0.03
	$\eta_g = 0.201$ (0.003)			

$g = 9$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	-4.90 (0.07)	-0.02 (0.01)	-0.02	0.01
$m = 2$	-0.90 (0.06)	-0.35 (0.04)	-0.33	0.29
$m = 3$	1.36 (0.06)	0.18 (0.03)	0.17	0.79
$m = 4$	-2.65 (0.10)	-0.19 (0.04)	-0.19	0.07
$m = 5$	-0.28 (0.06)	-0.91 (0.03)	-0.67	0.44
$m = 6$	-1.43 (0.07)	-0.60 (0.05)	-0.51	0.20
$m = 7$	0.49 (0.06)	-0.60 (0.04)	-0.52	0.62
$m = 8$	-4.11 (0.08)	-0.05 (0.01)	-0.05	0.02
$m = 9$	-2.77 (0.08)	-0.09 (0.02)	-0.09	0.06
$m = 10$	-8.48 (0.21)	0.00 (0.00)	0.00	0
$m = 11$	-2.25 (0.08)	0.01 (0.02)	0.01	0.10
$m = 12$	2.51 (0.08)	0.04 (0.03)	0.04	0.92
$m = 13$	-1.26 (0.07)	0.03 (0.04)	0.03	0.22
$m = 14$	-6.28 (0.04)	0.00 (0.00)	0.00	0
$m = 15$	-3.60 (0.08)	-0.05 (0.02)	-0.05	0.03
$m = 16$	-4.55 (0.06)	0.03 (0.01)	0.03	0.01
	$\eta_g = 0.094$ (0.003)			

$g = 10$

	b_{mg}	w_{mg}	w^*_{mg}	$\pi_{mg}(0)$
$m = 1$	-1.83 (0.09)	0.74 (0.08)	0.59	0.16
$m = 2$	1.52 (0.09)	0.29 (0.07)	0.28	0.82
$m = 3$	3.70 (0.10)	-0.06 (0.04)	-0.06	0.98
$m = 4$	0.36 (0.09)	0.97 (0.06)	0.70	0.58
$m = 5$	2.95 (0.12)	0.35 (0.06)	0.33	0.95
$m = 6$	1.05 (0.09)	0.94 (0.06)	0.68	0.71
$m = 7$	5.44 (0.09)	-0.06 (0.03)	-0.06	1
$m = 8$	0.16 (0.10)	-1.40 (0.05)	-0.81	0.53
$m = 9$	1.35 (0.10)	-1.29 (0.07)	-0.79	0.74
$m = 10$	-0.03 (0.10)	-1.11 (0.06)	-0.74	0.50
$m = 11$	1.86 (0.12)	-0.36 (0.09)	-0.34	0.86
$m = 12$	3.27 (0.15)	-0.02 (0.06)	-0.02	0.96
$m = 13$	1.44 (0.11)	0.19 (0.09)	0.19	0.81
$m = 14$	-1.88 (0.11)	0.30 (0.08)	0.28	0.14
$m = 15$	-1.33 (0.10)	0.18 (0.08)	0.18	0.21
$m = 16$	-3.66 (0.12)	0.10 (0.06)	0.10	0.03
	$\eta_g = 0.057$ (0.002)			

6.1.1 Analysis of the Groups in the Selected Model

From Table 4 it is possible to observe that Groups 7 and 8 consist mainly of the people that are able to do all the activities examined by the NLTCS survey. However, Group 8 has higher probability of being unable to do the activities. In particular, the probability of not being able to do heavy house work (activity 7) is much higher in Group 8 than in Group 7. In both groups, the estimated $\mathbf{w}_g$ values are zero, indicating that the local independence assumption is adequate in these groups. The fact that $\mathbf{w}_g$ is zero also indicates that there is no more variability in each outcome than can be described by a Bernoulli distribution.

In contrast, Group 6 consists mainly of disabled people. In addition, the $\mathbf{w}_g$ value is non-zero and the slope parameters are very negative for the activities of managing money (activity 14), taking medicine (activity 15) and telephoning (activity 16). The slope parameters for these activities indicate that there is considerable variability in the outcomes in these activities and that the outcomes are positively dependent (that is, those able/disabled in one activity will tend to be able/disabled in the others). It is worth noting that these activities are the least physically taxing of the instrumental activities of daily living, as recorded in these data.

Group 9 is characterized by people able to do the activities 1, 4, 8, 9, 10, 11, 13, 14, 15, 16 that correspond to eating, dressing, doing light house work, doing laundry, cooking, grocery shopping, travelling, managing money, taking medicine, telephoning, but these individuals are unable to get around inside and outside (activities 3 and 12). For the other activities (number 2, 5, 6, 7) there is quite a big variability in responses explained by the within group latent trait structure; the slope parameters for these activities are all negatively signed which again indicates that these activities exhibit strong positive dependence. These activities getting in/out of bed (activity 2), bathing (activity 5), using the toilet (activity 6) and doing heavy housework (activity 7) require mobility which may explain the positive dependence between the outcomes in these.

Group 1 consists mainly of people able to eat (activity 1), but unable to do heavy house work (activity 7), for the other activities there is quite a big variability explained by the within group latent trait structure, but the median individual in this group has a probability in excess of 80% to be disabled in activities 9, 10, 11, 12, 13. Activities 2, 3, 4, 5, 6, 10, 11, 12, 13, 14, 15, 16 have highly positive slope parameters, whereas activities 7, 8, 9 have moderately negative slopes. Thus there is positive dependence within these sets of activities but negative dependence across these groups.

Group 2 consists mainly of people able to get in/out of bed, and get around outside (activities 2,3), but unable to travel (activity 13), for the other activities there is quite a big variability

explained by the within group latent trait structure, but the probability of the median individual of this group being disabled is large (over the 88%) for the activities 11, 14 (grocery shopping and managing money), and small (less than the 10%) for the activities 1 and 6 (eating and using the toilet). Interestingly, the median individual tends to have lower probability of being disabled in the activities of daily living (ADLs) than in the instrumental activities of daily living (IADLs). The slopes for most activities have equal sign, so there is evidence of strong positive dependence between activities in this group. Thus people are either able or unable to do multiple activities in this group to a greater degree than independence would suggest.

Group 3 consists mainly of people able to do the activities 1, 4, 8, 10, 14, 15, 16, but unable to do the activities 3, 7, 11, 12, 13, for the other activities there is quite a big variability explained by the within group latent trait structure. The activities with large slope parameters involve physical movement and these again exhibit positive dependence in a similar manner to Group 9.

Group 4 is characterized mainly by people unable to do the activities of daily living 2, 3, 4, 5, 6, 7, but there is large variability in the responses explained by the within group latent trait structure. In contrast to Group 2, for this group the median individual tends to have a higher probability of being disabled in the activities of daily living (ADLs) than in the instrumental activities of daily living (IADLs). The signs of the large magnitude slope parameters are equal, so there is strong positive dependence within this group also.

Group 5 consists mainly of people able to do the activities 1, 2, 3, 4, 6, 8, 10, 15, 16, but unable to do heavy house work (activity 7), for the other activities there is quite a big variability explained by the within group latent trait structure. The large magnitude slope parameters are equal in sign, so there is strong positive dependence within these activities within this group also.

Group 10 consists mainly of people able to telephone (activity 16), but unable to do heavy house work and get around inside and outside (activities 7, 3 and 12), for the other activities there is quite a big variability explained by the within group latent trait structure, but the probability of being disabled for the median individual of this group is large (over the 80%) for the variables 2, 5, 11, 13, and quite small (between the 14% and the 21%) for the variables 1, 14, 15. In this group, activities 1, 2, 4, 5, 6, 14, 15 have strongly positive slope and activities 8, 9, 10, 11 have negative slopes. Thus, there is positive dependence within these activity groups and negative dependence across the activity groups. The second group of activities require a large amount of movement whereas the first group require less so.

The groups found in this analysis suggest that there is a wide range of experiences in terms of activities of daily living in the individuals included in the NLTCS data. The groups have a similar

structure to the one found by Erosheva et al. (2007) but their model allows individuals to have mixed membership across groups, so they suggest that slightly fewer groups (between 7 and 9) are needed.

6.2   U.S. Congressional Voting

The voting records of the U.S. House of Representatives Congressmen in the second session of the 98th congress (January 23rd, 1984–October 12th, 1984) are recorded in Congressional Quarterly Almanac (1984). During this session the House of Representatives recorded votes on 408 issues. Sixteen of the key issues voted on by the House are described in this publication (Congressional Quarterly Almanac, 1984, Pages 7-C to 11-C). The voting data for these sixteen issues was collated in Schlimmer (1987) and is publicly available in the UCI Machine Learning data repository (Asuncion and Newman, 2007).

The data set contains the votes of all 435 House of Representatives congressmen on the sixteen issues: each vote has a value of *yes* (voted for, paired for, and announced for), *no* (voted against, paired against, and announced against) or *undecided* (voted present, voted present to avoid conflict of interest, and did not vote or otherwise make a position known). It is also known if the voter is a *Democrat* or a *Republican* congressman. The issues voted on were: *(Q1)* Handicapped Infants, *(Q2)* Water Project Cost-Sharing, *(Q3)* Adoption of the Budget Resolution, *(Q4)* Physician Fee Freeze, *(Q5)* El Salvador Aid, *(Q6)* Religious Groups in Schools, *(Q7)* Anti-Satellite Test Ban, *(Q8)* Aid to Nicaraguan 'Contras', *(Q9)* MX Missile, *(Q10)* Immigration, *(Q11)* Synfuels Corporation Cutback, *(Q12)* Education Spending, *(Q13)* Superfund Right to Sue, *(Q14)* Crime, *(Q15)* Duty-Free Exports, *(Q16)* Export Administration Act/South Africa.

Each question has been coded using two binary variables a and b: the responses for the a variables are coded as $1 = yes/no$ and $0 = undecided$, and b variable are $1 = yes$, $0 = no/undecided$. The MLTA model was fitted to these data for $D = 0, 1, 2, 3$ and $G = 1, 2, \ldots, 5$. The log-likelihood for the fitted models are reported in Table 5.

The BIC and the BIC* (Table 6 and 7) agree in selecting the model with 4 groups, a bivariate latent trait and with slope parameters, $\mathbf{w}$, equal across groups as the best model.

6.2.1   Analysis of the Groups in the Selected Model

Table 8 shows a cross tabulation of group membership with party membership (*Republican* or *Democrat*) for the 435 congressmen. From this table it is clear that Group 4 consists mainly of

Table 5: Approximated log-likelihood for the models with $D = 0, 1, 2, 3$ and $G = 1, 2, \ldots, 5$

	$D=0$	$D=1$	**w** *fixed*	$D=2$	**w** *fixed*	$D=3$	**w** *fixed*
$G=1$	-6109.61	-4789.10		-4565.47		-4468.45	
$G=2$	-4888.64	-4533.89	-4741.79	-4364.65	-4492.78	-4317.43	-4383.82
$G=3$	-4699.47	-4332.34	-4580.95	-4245.82	-4417.75	-4174.52	-4340.71
$G=4$	-4613.10	-4212.98	-4453.24	-4141.93	-4260.51	-4074.31	-4378.15
$G=5$	-4533.50	-4149.19	-4378.84	-4037.96	-4241.45	-3956.23	-4236.31

Table 6: The estimated BIC for the models with $D = 0, 1, 2, 3$ and $G = 1, 2, \ldots, 5$

	$D=0$	$D=1$	**w** *fixed*	$D=2$	**w** *fixed*	$D=3$	**w** *fixed*
$G=1$	12413.64	9967.02		9708.10		9696.32	
$G=2$	10172.18	9851.49	10072.90	9889.69	9763.21	10159.77	9727.54
$G=3$	9994.32	9843.29	9951.70	10235.27	9813.63	10639.44	9841.81
$G=4$	10022.07	9999.47	9896.76	10610.72	**9699.65**	11204.53	10117.18
$G=5$	10063.36	10266.79	9948.44	10986.01	9862.00	11733.84	10033.99

Republican congressmen, Group 1 and 3 consist mainly of *Democrat* congressmen and Group 2 is a small group consisting of members of both parties. Interestingly, even though the Democrat congressmen are divided into two voting blocs, these do not correspond to the Northern Democrats and Southern Democrats which were perceived to be voting blocs within the party (Table 9). However, of the eleven *Democrat* congressmen who are in Group 4, nine are Southern Democrats, suggesting that they vote in a similar way to the *Republican* congressmen. It is worth emphasizing that the data under consideration only records voting on sixteen key issues and that more voting blocs may be revealed if voting on all issues were examined.

Table 10 gives the parameter estimate b_{mg} and median probability $\pi_{mg}(\mathbf{0})$ for each of the groups. This table reveals that Group 2 consists of congressmen who have a high probability of being *undecided* on many of the issues.

The probabilities of a positive response for the a variables for the median individuals in Groups 1, 3 and 4 are always very large (over the 0.90) with two exceptions in Group 1, for the variable number 3 where $\pi_{31}(\mathbf{0}) = 0.82$ and for variable number 31 where $\pi_{311}(\mathbf{0}) = 0$. So, the majority of congressmen in the Groups 1, 3 and 4 voted on most issues, but with a slightly lower voting rate in Group 1 on the Water Project Cost-Sharing and a very low voting rate on the Export Administration Act/South Africa. As a result of these high voting rates, most of the probabilities

Table 7: The estimated BIC* for the models with $D = 0, 1, 2, 3$ and $G = 1, 2, \ldots, 5$

	$D = 0$	$D = 1$		$D = 2$		$D = 3$	
			w *fixed*		**w** *fixed*		**w** *fixed*
$G = 1$	12413.64	9967.02		9708.10		9696.32	
$G = 2$	10127.64	9761.98	10028.51	9746.68	9718.82	9986.49	9681.51
$G = 3$	9883.15	9614.18	9801.00	9871.50	9698.52	10204.99	9729.45
$G = 4$	9834.05	9600.57	9666.16	10017.47	**9464.28**	10404.27	9920.84
$G = 5$	9787.05	9711.72	9615.51	10128.07	9572.70	10531.05	9741.57

Table 8: Classification table for party in the selected model ($G = 4$, $D = 2$, **w** *fixed*)

	$g = 1$	$g = 2$	$g = 3$	$g = 4$
Democrat	65	8	183	11
Republican	2	3	18	145

given for b variables (*yes* versus *no/undecided*) in these groups can be interpreted as a voting *yes* versus *no* probability.

From Table 10 and Figure 2 it is possible to observe that the responses for the median individual in the mainly *Republican* group (Group 4) are the opposite to the ones given by the median individuals in the mainly *Democrat* groups (Groups 1 and 3) for the majority of the issues.

In fact the median individual in the *Republican* group (Group 4) has high probabilities to give positive responses for the variables $4b, 5b, 6b, 12b, 13b, 14b$, which are concerned with the Physician Fee Freeze, the El Salvador Aid, the Religious Groups in Schools, the Education Spending, the Superfund Right to Sue and the Crime; the *Democrat* groups (Groups 1 and 3) have low probabilities of giving these issues a positive response. Additionally, the *Democrat* groups have high probability of positive responses for the variables $3b, 7b, 8b, 9b$, which are concerned with the Adoption of the Budget Resolution, the Anti-Satellite Test Ban, the Aid to the Nicaraguan 'Contras' and the MX Missile, whereas the *Republican* groups have low probabilities for these issues.

The choice of a model with the **w** common in all groups suggests the latent trait has the same effect in all groups. The estimated posterior mean of the latent variable $\mathbf{y}_n$ conditional on $z_{ng} = 1$ was estimated for each voter and for each group and these are shown in Figures 1–4 of the supplementary material. The posterior mean is plotted a number of times in these figures to show the impact of the latent variable on outcome variables. The variation in the positions of the congressmen in these plots shows how the voting is influenced by the latent trait in addition to the baseline median probability of voting as characterized by the intercept parameters **b**. It is

Table 9: Classification table for location and party in the selected model

	$g = 1$	$g = 2$	$g = 3$	$g = 4$
Northern Democrat	48	7	121	2
Southern Democrat	17	1	62	9
Northern Republican	2	2	13	111
Southern Republican	0	1	5	34

clear from the plots that the latent trait can introduce significant variation in the outcome variable within a group if the slope parameter $\mathbf{w}_g$ has large magnitude.

From the analysis of the *lift* (Tables 2–5 of the supplementary material) it is possible to observe that: within Group 1 and Group 3 there is a strong positive dependence within the set of variables $(4b, 5b, 6b, 12b, 13b, 14b)$; this is shown by *lift* values in excess of one. Within Group 1 there is also a negative dependence between this set of variables and the set $(16a, 16b)$, within which there is a positive dependence; this is evidenced by the *lift* values well below one. Within Group 4 there is a positive dependence within the set of variables $(3b, 7b, 8b, 9b, 15b)$, and a negative dependence between this set of variables and the variable $2b$. In Group 2 there is a strong positive dependence within the sets of variables $(4b, 5b, 6b, 12b, 13b, 14b)$, $(2b, 4b, 5b, 6b)$ and $(15b, 16a, 16b)$, and a negative dependence between the sets of variables $(2b, 4b, 5b, 6b)$ and $(3b, 7b, 8b, 9b, 15b, 16a, 16b)$, and between the sets variables $(12b, 13b, 14b)$ and $(15b, 16a, 16b)$. These *lift* values show that the model is accounting for dependence within the groups which may be missed in a latent class analysis of these data.

7 Conclusions

The mixture of latent trait analyzers model provides a powerful methodology for model-based clustering for categorical data. It provides a suitable alternative to LCA for clustering binary data and it can be fitted in a computationally efficient manner using a variational EM algorithm. In addition, the model includes LCA as a special case so it is also applicable when the LCA modeling assumptions are appropriate. The continuous latent trait in model also facilitates investigating the dependence between variables, thus offering a combination of the properties offered by the Latent Class Analysis model and the Latent Trait Analysis model. The model parameters are interpretable and provide a characterization of the within cluster structure.

The variational EM algorithm proposed for model fitting gives an efficient mechanism for model

Table 10: Parameter estimates (*standard errors*) for the best model selected ($G = 4$, $D = 2$, **w** *fixed*). The variable coded as a records the congress person voting *yes/no* versus *undecided* and the vote coded as b records the congress person voting *yes* versus *no/undecided*. In both cases the median probability gives the probability of the first outcome.

		$g=1$		$g=2$		$g=3$		$g=4$		$d=1$		$d=2$	
		b_{mg}	$\pi_{mg}(\mathbf{0})$	b_{mg}	$\pi_{mg}(\mathbf{0})$	b_{mg}	$\pi_{mg}(\mathbf{0})$	b_{mg}	$\pi_{mg}(\mathbf{0})$	w_{dm}	w^*_{dm}	w_{dm}	w^*_{dm}
Q1	a	4.23 (1.99)	0.98	-0.98 (0.89)	0.28	4.23 (0.56)	0.99	9.67 (0.25)	1	-0.24 (0.23)	-0.23	0.14 (0.08)	0.14
	b	1.10 (0.30)	0.73	-2.49 (2.22)	0.09	0.24 (0.16)	0.55	-1.74 (0.24)	0.16	-0.54 (0.13)	-0.44	0.45 (0.14)	0.37
Q2	a	1.55 (0.33)	0.82	-2.45 (2.26)	0.09	3.04 (0.32)	0.95	2.23 (0.28)	0.90	0.35 (0.19)	0.31	0.30 (0.12)	0.27
	b	-0.35 (0.29)	0.43	-24.80 (1.81)	0	0.02 (0.18)	0.50	-0.38 (0.20)	0.43	0.90 (0.16)	0.57	0.81 (0.16)	0.52
Q3	a	10.39 (0.31)	1	-0.17 (0.88)	0.46	4.21 (0.63)	0.98	4.52 (0.81)	0.99	-0.39 (0.22)	-0.36	0.02 (0.07)	0.02
	b	3.35 (0.60)	0.94	-0.72 (1.16)	0.37	2.08 (0.28)	0.83	-2.43 (0.32)	0.12	-1.21 (0.19)	-0.77	0.06 (0.12)	0.04
Q4	a	11.72 (0.41)	1	0.21 (0.62)	0.55	3.67 (0.36)	0.97	11.17 (0.36)	1	-0.56 (0.16)	-0.49	-0.08 (0.10)	-0.07
	b	-3.84 (0.74)	0.04	-15.91 (1.12)	0	-2.55 (0.25)	0.13	3.58 (0.53)	0.95	1.33 (0.19)	0.80	-0.11 (0.12)	-0.07
Q5	a	3.11 (0.69)	0.95	0.19 (0.84)	0.54	3.40 (0.37)	0.97	9.30 (0.35)	1	-0.36 (0.22)	-0.34	-0.09 (0.10)	-0.09
	b	-3.69 (0.64)	0.11	-26.22 (2.14)	0	-1.93 (0.25)	0.26	5.97 (0.78)	0.97	2.41 (0.20)	0.92	0.04 (0.08)	0.01
Q6	a	3.50 (0.84)	0.97	0.18 (0.86)	0.55	3.94 (0.49)	0.98	10.26 (0.39)	1	0.01 (0.24)	0.01	-0.24 (0.11)	-0.23
	b	-0.88 (0.38)	0.36	-1.98 (1.36)	0.27	0.30 (0.22)	0.52	3.44 (0.43)	0.90	2.06 (0.20)	0.88	-0.48 (0.14)	-0.21
Q7	a	3.46 (0.74)	0.95	0.76 (0.88)	0.64	4.20 (0.38)	0.97	4.99 (2.53)	0.99	-1.00 (0.23)	-0.71	-0.11 (0.12)	-0.08
	b	3.44 (0.59)	0.89	0.61 (1.15)	0.55	1.43 (0.26)	0.70	-2.07 (0.27)	0.21	-2.22 (0.19)	-0.91	-0.09 (0.10)	-0.04
Q8	a	16.20 (1.09)	1	1.33 (0.93)	0.73	3.95 (0.42)	0.96	3.84 (0.55)	0.97	-1.04 (0.30)	-0.71	-0.25 (0.15)	-0.17
	b	30.61 (2.79)	1	1.57 (1.23)	0.64	2.19 (0.30)	0.75	-3.73 (0.38)	0.11	-2.55 (0.21)	-0.93	-0.14 (0.11)	-0.05
Q9	a	2.25 (0.41)	0.90	1.09 (0.90)	0.73	3.09 (0.33)	0.95	4.64 (1.83)	0.99	-0.55 (0.14)	-0.47	-0.19 (0.11)	-0.16
	b	1.59 (0.41)	0.74	1.21 (1.28)	0.64	1.07 (0.23)	0.67	-3.26 (0.37)	0.10	-1.94 (0.18)	-0.88	-0.28 (0.13)	-0.13
Q10	a	3.51 (0.95)	0.97	0.58 (0.83)	0.64	10.46 (0.28)	1	5.11 (2.29)	0.99	-0.29 (0.23)	-0.27	-0.16 (0.12)	-0.15
	b	-0.58 (0.28)	0.37	-1.17 (1.36)	0.27	0.10 (0.16)	0.52	0.14 (0.18)	0.53	-0.13 (0.13)	-0.11	-0.74 (0.16)	-0.59
Q11	a	3.54 (0.75)	0.97	-1.03 (1.36)	0.27	3.41 (0.34)	0.96	3.74 (0.48)	0.98	-0.40 (0.25)	-0.37	-0.15 (0.14)	-0.13
	b	-0.37 (0.26)	0.41	-2.43 (3.37)	0.09	-0.05 (0.16)	0.49	-1.80 (0.29)	0.15	0.42 (0.18)	0.38	0.20 (0.22)	0.18
Q12	a	3.56 (0.91)	0.97	-0.20 (0.97)	0.45	2.44 (0.29)	0.91	3.44 (0.80)	0.97	-0.38 (0.23)	-0.34	-0.29 (0.18)	-0.26
	b	-2.38 (0.42)	0.12	-1.86 (1.17)	0.18	-2.32 (0.25)	0.13	2.18 (0.39)	0.87	0.96 (0.18)	0.66	-0.45 (0.16)	-0.31
Q13	a	2.87 (0.53)	0.94	0.63 (0.85)	0.64	2.71 (0.27)	0.93	4.28 (0.55)	0.98	-0.55 (0.20)	-0.48	-0.15 (0.15)	-0.13
	b	-2.32 (0.44)	0.13	-1.32 (2.06)	0.27	-0.98 (0.19)	0.31	2.20 (0.30)	0.87	1.08 (0.18)	0.73	-0.18 (0.20)	-0.12
Q14	a	3.71 (0.67)	0.97	0.68 (0.94)	0.64	3.13 (0.28)	0.95	9.54 (1.08)	1	-0.69 (0.24)	-0.55	-0.26 (0.16)	-0.21
	b	-1.70 (0.36)	0.22	-1.38 (0.97)	0.27	-0.68 (0.20)	0.37	11.16 (3.25)	1	1.16 (0.19)	0.70	-0.65 (0.15)	-0.39
Q15	a	2.56 (0.47)	0.92	-0.19 (0.96)	0.46	2.98 (0.33)	0.95	3.19 (0.45)	0.96	-0.37 (0.24)	-0.35	-0.14 (0.15)	-0.13
	b	0.81 (0.31)	0.66	-1.94 (1.50)	0.18	0.31 (0.18)	0.57	-2.80 (0.35)	0.08	-1.08 (0.16)	-0.73	0.07 (0.15)	0.05
Q16	a	-26.79 (2.38)	0	-3.04 (2.13)	0.18	4.17 (0.35)	0.91	3.78 (0.33)	0.93	-2.18 (0.17)	-0.91	0.01 (0.12)	0.00
	b	-33.71 (3.05)	0	-3.69 (2.54)	0.18	3.76 (0.39)	0.85	0.99 (0.28)	0.61	-2.74 (0.22)	-0.94	0.03 (0.14)	0.01
		$\eta_1 = 0.152$	(0.017)	$\eta_2 = 0.025$	(0.009)	$\eta_3 = 0.465$	(0.029)	$\eta_4 = 0.358$	(0.027)				
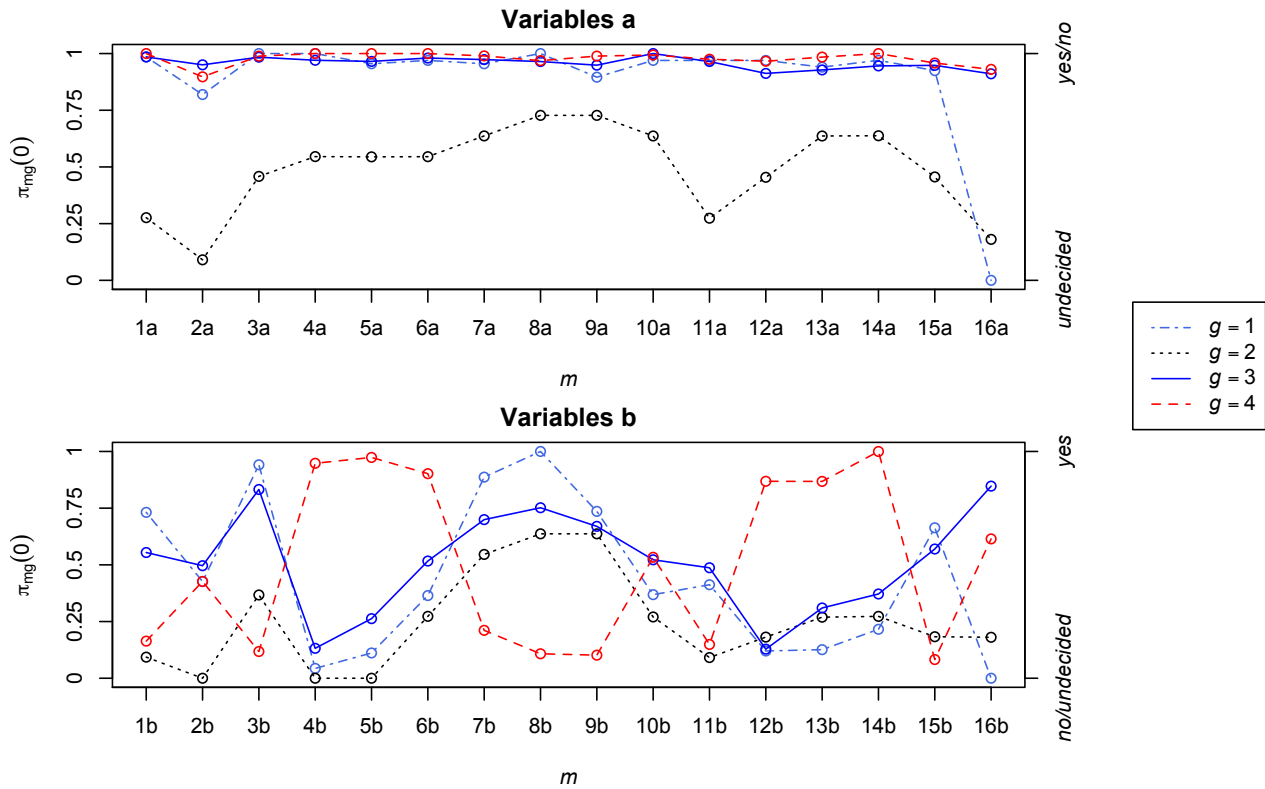
Figure 2: The median probabilities, $\pi_{mg}(\mathbf{0})$, for the selected model ($G = 4$, $D = 2$, $\mathbf{w}$ *fixed*). The variable coded as a records the congress person voting *yes/no* versus *undecided* and the vote coded as b records the congress person voting *yes* versus *no/undecided*. In both cases the median probability gives the probability of the first outcome.

fitting, thus making the method easily applicable.

The model provides an extra model-based clustering framework for categorical data with a similar structure to the continuous data methods of mclust (Fraley and Raftery, 2002), MIXMOD (Biernacki et al., 2006), EMMIX (McLachlan and Peel, 1999), pgmm (McNicholas and Murphy, 2008) and t-EIGEN(Andrews and McNicholas, 2012).

The MLTA framework as proposed for the automatic selection using BIC (and a variant) of the number of clusters (G) and the latent trait dimensionality (D) required to appropriately model dependence within clusters. Alternative model selection techniques that are not based on a penalized likelihood criterion could offer a more computationally efficient manner for selecting G and D.

The excellent clustering behavior of this method has been shown by two applications on the National Long Term Care Survey (NLTCS) and U.S. Congressional Voting data sets. In both cases, the model found groups that were intuitive in their interpretation and fewer groups were found than the LCA model suggests.

Acknowledgements

References

ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. 9th ed. Dover, New York.

ALLMAN, E. S., MATIAS, C. and RHODES, J. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, **37** 3099–3132.

ANDREWS, J. L. and MCNICHOLAS, P. D. (2012). Model-based clustering, classification, and discriminant analysis via mixtures of multivariate *t*-distributions: The *t*EIGEN family. *Statistics and Computing*, **22** 1021–1029.

ASUNCION, A. and NEWMAN, D. J. (2007). UCI machine learning repository. URL `http://www.ics.uci.edu/~mlearn/MLRepository.html`.

BAEK, J., McLACHLAN, G. and FLACK, L. (2010). Mixtures of factor analyzers with common factor loadings: applications to the clustering and visualisation of high-dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32** 1298–1309.

BARTHOLOMEW, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, **42** pp. 293–321. URL `http://www.jstor.org/stable/2985165`.

BARTHOLOMEW, D. J., KNOTT, M. and MOUSTAKI, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach.* 3rd ed. Wiley.

BARTHOLOMEW, D. J., STEELE, F., MOUSTAKI, I. and GALBRAITH, J. (2002). *The Analysis and Interpretation of Multivariate Data for Social Scientists.* Chapman & Hall Ltd.

BIERNACKI, C., CELEUX, G., GOVAERT, G. and LANGROGNET, F. (2006). Model-based cluster and discriminant analysis with the MIXMOD software. *Computational Statistics and Data Analysis*, **51** 587–600.

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning.* Information Science and Statistics, Springer, New York.

BOCK, R. D. and AITKIN, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, **46** 443–459.

BRIN, S., MOTWANI, R., ULLMAN, J. D. and TSUR, S. (1997). Dynamic itemset counting and implication rules for market basket data. *SIGMOD Record*, **26** 255–264.

CELEUX, G. and GOVAERT, G. (1995). Gaussian parsimonious clustering models. *Pattern Recognition*, **28** 781 – 793.

CONGRESSIONAL QUARTERLY ALMANAC (1984). *98th congress, 2nd session*, volume XL ed.

DEAN, N. and RAFTERY, A. (2010). Latent class analysis variable selection. *Annals of the Institute of Statistical Mathematics*, **62** 11–35.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood for incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39** 1–38.

EFRON, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, **68** 589–599.

EROSHEVA, E. A. (2002). *Grade of Membership and Latent Structure Models with Application to Disability Survey Data.* Ph.D. thesis, Department of Statistics, Carnegie Mellon University.

EROSHEVA, E. A. (2003). Bayesian estimation of the grade of membership model. In *Bayesian Statistics 7* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.). Oxford, 501–510.

EROSHEVA, E. A. (2004). Partial membership models with application to disability survey data. In *Statistical Data Mining and Knowledge Discovery* (H. Bozdogan, ed.). CRC Press, 117–134.

EROSHEVA, E. A., FIENBERG, S. E. and JOUTARD, C. (2007). Describing disability through individual-level mixture models for multivariate binary data. *Annals of Applied Statistics*, **1** 502–537.

FIENBERG, S. E., HERSH, P., RINALDO, A. and ZHOU, Y. (2009). Maximum likelihood estimation in latent class models for contingency tables. In *Algebraic and Geometric Methods in Statistics* (P. Gibilisco, E. Riccomagno, M. Rogantin and H. Wynn, eds.). Cambridge University Press, 31–66.

FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, **97** 611–612.

GHAHRAMANI, Z. and HINTON, G. E. (1997). The EM algorithm for mixtures of factor analyzers. Tech. Rep. CRG-TR-96-1, University Of Toronto, Toronto.

GOODMAN, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, **61** pp. 215–231. URL http://www.jstor.org/stable/2334349.

HADGU, A. and QU, Y. (1998). A biomedical application of latent class models with random effects. *Applied Statistics*, **47** 603–616.

JAAKKOLA, T. S. and JORDAN, M. I. (1996). A variational approach to Bayesian logistic regression models and their extensions. In *Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics.*

KARLIS, D. and SANTOURIAN, A. (2008). Model-based clustering with non-elliptically contoured distributions. *Statistics and Computing*, **19** 73–83.

LIN, T. I. (2010). Robust mixture modeling using multivariate skew *t* distributions. *Statistics and Computing*, **20** 343–356.

LIN, T. I., LEE, J. C. and YEN, S. Y. (2007). Finite mixture modeling using the skew normal distribution. *Statistica Sinica*, **17** 909–927.

MCLACHLAN, G. and PEEL, D. (1999). The EMMIX algorithm for the fitting of normal and t-components. *Journal of Statistical Software*, **4** 1–14. URL http://www.jstatsoft.org/v04/i02.

MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York.

MCLACHLAN, G., PEEL, D. and BEAN, R. (2003). Modelling high-dimensional data by mixtures of factor analyzers. *Computational Statistics & Data Analysis*, **41** 379–388.

MCNICHOLAS, P. D. and MURPHY, T. B. (2008). Parsimonious Gaussian mixture models. *Statistics and Computing*, **18** 285–296.

MCNICHOLAS, P. D. and MURPHY, T. B. (2010). Model-based clustering of microarray expression data via latent Gaussian mixture models. *Bioinformatics*, **26** 2705–2712.

MUTHÉN, B. (2001). Latent variable mixture modeling. In *New developments and techniques in structural equation modeling* (G. A. Marcoulides and R. E. Schumacker, eds.). Lawrence Erlbaum Associates, 1–33.

PAULER, D. K. (1998). The Schwarz criterion and related methods for normal linear models. *Biometrika*, **85** 13–27.

QU, Y., TAN, M. and KUTNER, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, **52** 797–810.

RAFTERY, A. E., NEWTON, M. A., SATAGOPAN, J. M. and KRIVITSKY, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In *Bayesian statistics 8*. Oxford University Press, Oxford, 1–45.

RASCH, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.* Nielsen & Lydiche, Oxford, England.

ROST, J. (1990). Rasch models in latent classes: an integration of two approaches to item analysis. *Applied Psychological Measurement*, **14** 271–282.

ROST, J. and VON DAVIER, M. (1995). Mixture distribution Rasch models. In *Rasch Models: Foundations, Recent Developments, and Applications* (G. H. Fischer and I. W. Molenaar, eds.). Springer Verlag, New York, 257–268.

SAMMEL, M. D., RYAN, L. M. and LEGLER, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B, Methodological*, **59** 667–678.

SCHLIMMER, J. C. (1987). *Concept Acquisition Through Representational Adjustment.* Ph.D. thesis, Department of Information and Computer Science, University of California, Irvine.

SCHWARZ, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6** 461–464.

STEELE, R. J. (2002). *Practical Importance Sampling Methods for Finite Mixture Models and Multiple Imputation.* Ph.D. thesis, University of Washington.

TIPPING, M. E. (1999). Probabilistic visualisation of high-dimensional binary data. In *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems 11*. MIT Press, Cambridge, MA, USA, 592–598.

UEBERSAX, J. S. (1999). Probit latent class analysis with dichotomous or ordered category measures: Conditional independence/dependence models. *Applied Psychological Measurement*, **23** 283–297.

VERMUNT, J. (2007). Multilevel mixture item response theory models: An application in education testing. In *Proceedings of the 56th session of the International Statistical Institute, Lisbon, Portugal*. International Statistical Institute, Voorburg, Netherlands.

VERMUNT, J. and MAGIDSON, J. (2005). Factor analysis with categorical indicators: a comparison between traditional and latent class approaches. In *New developments in categorical data analysis for the social and behavioral sciences* (A. V. der Ark, M. A. Croon and K. Sijtsma, eds.). Lawrence Erlbaum Associates, 41–62.

Vermunt, J. and Magidson, J. (2008). *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module.* Belmont, MA, Statistical Innovations Inc.

von Davier, M., Rost, J. and Carstensen, C. H. (2007). Introduction: Extending the Rasch model. In *Multivariate and Mixture Distribution Rasch Models* (M. von Davier and C. H. Carstensen, eds.). Springer Verlag, New York, 1–12.

von Davier, M. and Yamamoto, K. (2007). Mixture distribution and HYBRID Rasch models. In *Multivariate and Mixture Distribution Rasch Models* (M. von Davier and C. H. Carstensen, eds.). Springer Verlag, New York, 99–115.

Isabella Gollini[†]   Thomas Brendan Murphy[⋆]

[†]National Centre for Geocomputation, National University of Ireland Maynooth, Ireland

[⋆]School of Mathematical Sciences, University College Dublin, Ireland

1 Parameter Estimation in the Parsimonious Model

It is necessary to use an EM algorithm to estimate the parameters in the parsimonious model presented in Section 3.1 of the main paper. The mathematical calculations required to derive the EM algorithm for the parsimonious model beyond those needed for the MLTA model are contained herein.

We define $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$, $\mathbf{Z} = (\mathbf{z}_{11}, \ldots, \mathbf{z}_{NG})$ and $\mathbf{\Xi} = (\boldsymbol{\xi}_{11}, \ldots, \boldsymbol{\xi}_{NG})$.

$$\log\left(\tilde{p}\left(\mathbf{X}|\mathbf{Y},\mathbf{Z},\mathbf{\Xi}\right)\right) = \sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\log\left(\prod_{m=1}^{M}\tilde{p}\left(x_{nm}|\mathbf{y}_n, z_{ng}=1, \xi_{nmg}\right)\right)$$

$$= \sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\lambda(\xi_{nmg})\left[\mathbf{w}_m^T\mathbf{y}_n\mathbf{y}_n^T\mathbf{w}_m + 2b_{mg}\mathbf{w}_m^T\mathbf{y}_n + b_{mg}^2\right] + \sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\left(x_{nm} - \frac{1}{2}\right)\left(\mathbf{w}_m^T\mathbf{y}_n + b_{mg}\right) + \text{const}_{\hat{\mathbf{w}}_m}$$

where $\text{const}_{\hat{\mathbf{w}}_m}$ are all the terms not involving $\hat{\mathbf{w}}_m$.

E-step

$$\mathbb{E}\left[\log\left(\tilde{p}\left(\mathbf{X}|\mathbf{Y},\mathbf{Z},\mathbf{\Xi}\right)\right)\right] = \sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\lambda(\xi_{nmg})\left[\mathbf{w}_m^T\mathbb{E}\left[\mathbf{y}_n\mathbf{y}_n^T\right]_g\mathbf{w}_m + 2b_{mg}\mathbf{w}_m^T\mathbb{E}[\mathbf{y}_n]_g + b_{mg}^2\right]$$

$$+ \sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\left[\left(x_{nm} - \frac{1}{2}\right)\left(\mathbf{w}_m^T\mathbb{E}[\mathbf{y}_n]_g + b_{mg}\right)\right] + \text{const}_{\hat{\mathbf{w}}_m}$$

$$= \mathbf{w}_m^T\left[\sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\lambda(\xi_{nmg})\mathbb{E}\left[\mathbf{y}_n\mathbf{y}_n^T\right]_g\right]\mathbf{w}_m + 2\mathbf{w}_m^T\left[\sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\lambda(\xi_{nmg})\boldsymbol{\mu}_{ng}b_{mg}\right] + \sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\lambda(\xi_{nmg})b_{mg}^2$$

$$+ \mathbf{w}_m^T\left[\sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\left(x_{nm} - \frac{1}{2}\right)\boldsymbol{\mu}_{ng}\right] + \sum_{g=1}^{G}\sum_{n=1}^{N} z_{ng}\left(x_{nm} - \frac{1}{2}\right)b_{mg} + \text{const}_{\hat{\mathbf{w}}_m}$$

$$= \hat{\mathbf{w}}_m^T\mathbf{K}_m\hat{\mathbf{w}}_m + \hat{\mathbf{w}}_m^T\boldsymbol{\gamma}_m + \text{const}_{\hat{\mathbf{w}}_m}$$

where

$$\hat{\mathbf{w}}_m = (\mathbf{w}_m^T, b_{m1}, \ldots, b_{mG})^T, \quad \boldsymbol{\gamma}_m = \begin{bmatrix} \sum_{g=1}^G \sum_{n=1}^N z_{ng} \left(x_{nm} - \frac{1}{2} \right) \boldsymbol{\mu}_{ng} \\ \sum_{n=1}^N z_{n1} \left(x_{nm} - \frac{1}{2} \right) \\ \vdots \\ \sum_{n=1}^N z_{nG} \left(x_{nm} - \frac{1}{2} \right) \end{bmatrix}$$

and,

$$\mathbf{K}_m = \begin{bmatrix} \sum_{g=1}^G \sum_{n=1}^N z_{ng} \lambda(\xi_{nmg}) \mathbb{E}\left[\mathbf{y}_n \mathbf{y}_n^T \right]_g & \sum_{n=1}^N z_{n1} \lambda(\xi_{nm1}) \boldsymbol{\mu}_{n1} & \cdots & \sum_{n=1}^N z_{nG} \lambda(\xi_{nmG}) \boldsymbol{\mu}_{nG} \\ \sum_{n=1}^N z_{n1} \lambda(\xi_{nm1}) \boldsymbol{\mu}_{n1}^T & \sum_{n=1}^N z_{n1} \lambda(\xi_{nm1}) & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ \sum_{n=1}^N z_{nG} \lambda(\xi_{nmG}) \boldsymbol{\mu}_{nG}^T & 0 & 0 & \sum_{n=1}^N z_{nG} \lambda(\xi_{nmG}) \end{bmatrix}$$

M-step

$$\hat{\mathbf{w}}_m = \arg\max_{\hat{\mathbf{w}}_m} \left(\mathbb{E}\left[\log \left(\tilde{p}\left(\mathbf{X} | \mathbf{Y}, \mathbf{Z}, \boldsymbol{\Xi} \right) \right) \right] \right)$$

$$\frac{\partial \mathbb{E}\left[\log \left(\tilde{p}\left(\mathbf{X} | \mathbf{Y}, \mathbf{Z}, \boldsymbol{\Xi} \right) \right) \right]}{\partial \hat{\mathbf{w}}_m} = 2\mathbf{K}_m \hat{\mathbf{w}}_m + \boldsymbol{\gamma}_m = 0$$

therefore,

$$\hat{\mathbf{w}}_m = - \left[2\mathbf{K}_m \right]^{-1} \boldsymbol{\gamma}_m$$

2 Additional Results

2.1 National Long Term Care Survey (NLTCS) Example

Table 1 presents the results mentioned in the example regarding the National Long Term Care Survey (NLTCS) (Section 6.1 of the main paper). It shows the comparison of the observed and the expected frequencies for the response patterns with more than 100 observations under the best model selected ($G = 10$, $D = 1$).

Table 1: Expected frequencies for the response patterns with more than 100 observations for the selected model

Response Pattern	Observed	Expected
0000000000000000	3853	3851
0000100000000000	216	225
0000001000000000	1107	1028
0000101000000000	188	228
0000001000100000	122	120
0000000000010000	351	351
0010000000010000	206	141
0000001000010000	303	312
0010001000010000	182	168
0000101000010000	108	98
0010101000010000	106	128
0000000000001000	195	200
0000001000001000	198	224
0000001000101000	196	162
0000001000011000	123	110
0000001000111000	176	176
0010001000111000	120	98
0000101000111000	101	111
0111111111111000	102	64
1111111111111010	107	88
0111111111111110	104	125
1111111111111110	164	218
0111111111111111	153	170
1111111111111111	660	506

2.2 U.S. Congressional Voting Example

This section reports some details of the results for the best model selected ($G = 4$, $D = 2$, $\mathbf{w}$ *fixed*) for the U.S. Congressional Voting Example mentioned in the paper (Section 6.2).

We show the plots (Figures 1–4), mentioned in Section 6.2.1, of the estimated expectations $\boldsymbol{\mu}_{ng}$ of the latent variable $\mathbf{y}_n$. We have a plot for each group, and only the observations with the maximum a posteriori probability to belong to that group are shown.

In the variables coded as a the green circles correspond to the congressmen voting *yes/no* and the yellow triangles to the people *undecided*. In the variables coded as b the green circles

3

correspond to the congressmen voting *yes* and the yellow triangles to the people *no/undecided*.

In the next pages are also reported Tables 2–5 showing the estimates of the *lift* mentioned in the paper (end of Section 6.2.1).
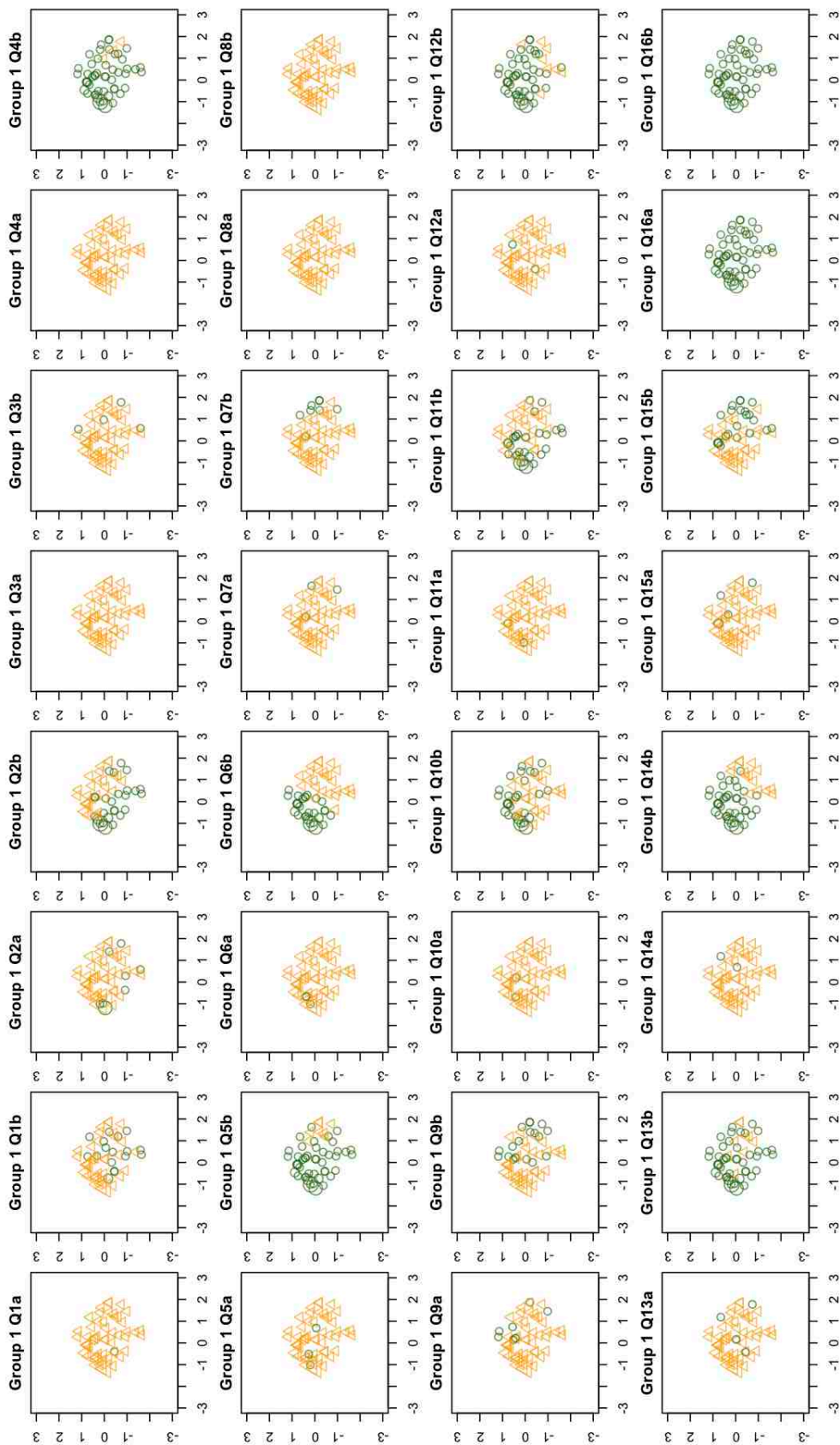
Figure 1: Plot of the $\boldsymbol{\mu}_{n1}$ for the selected model ($G = 4, D = 2, \mathbf{w} fixed$) in the U.S. congressional voting example.
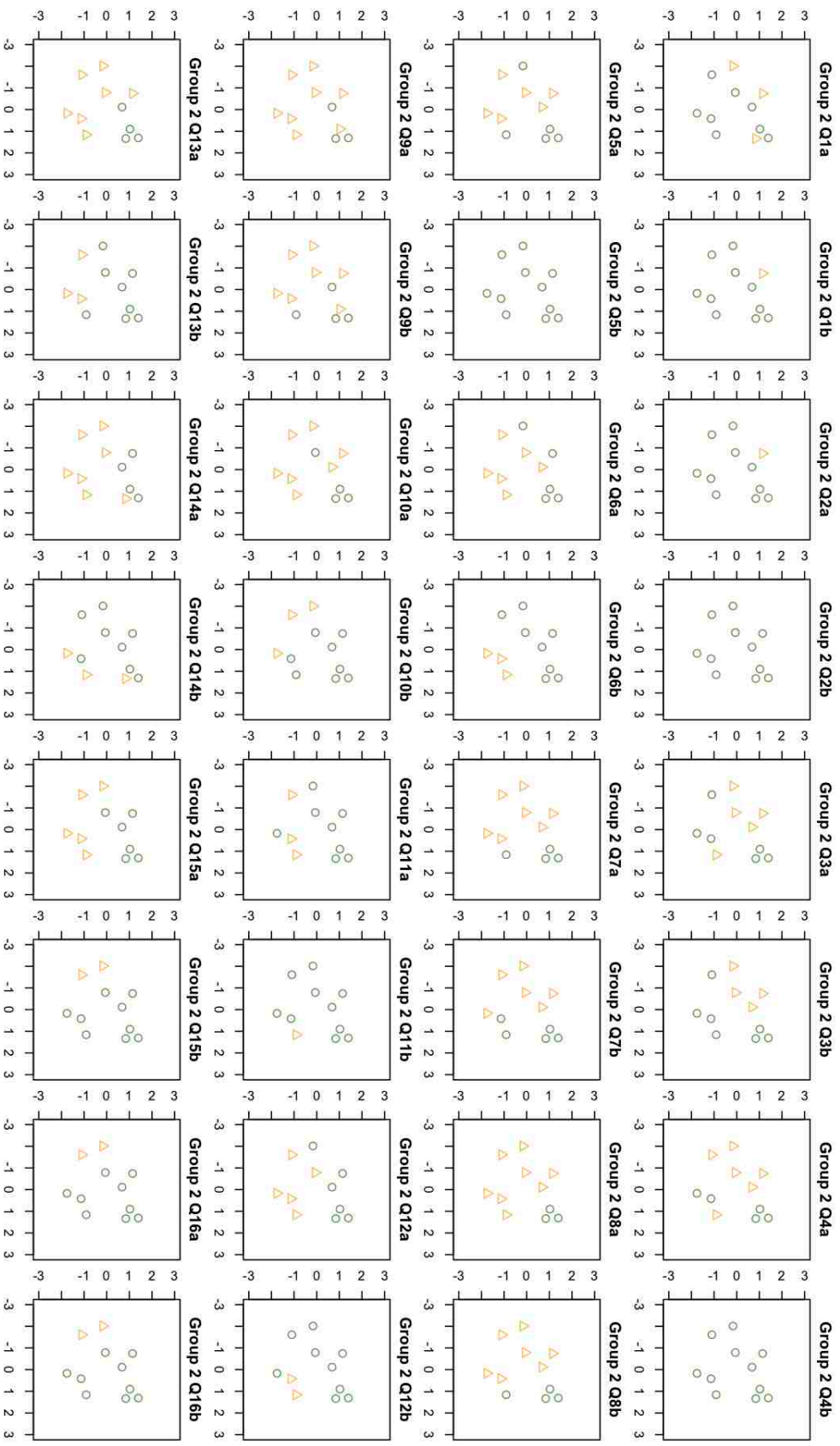
Figure 2: Plot of the $\boldsymbol{\mu}_{n2}$ for the selected model ($G = 4, D = 2, \mathbf{w}\,fixed$) in the U.S. congressional voting example.

6
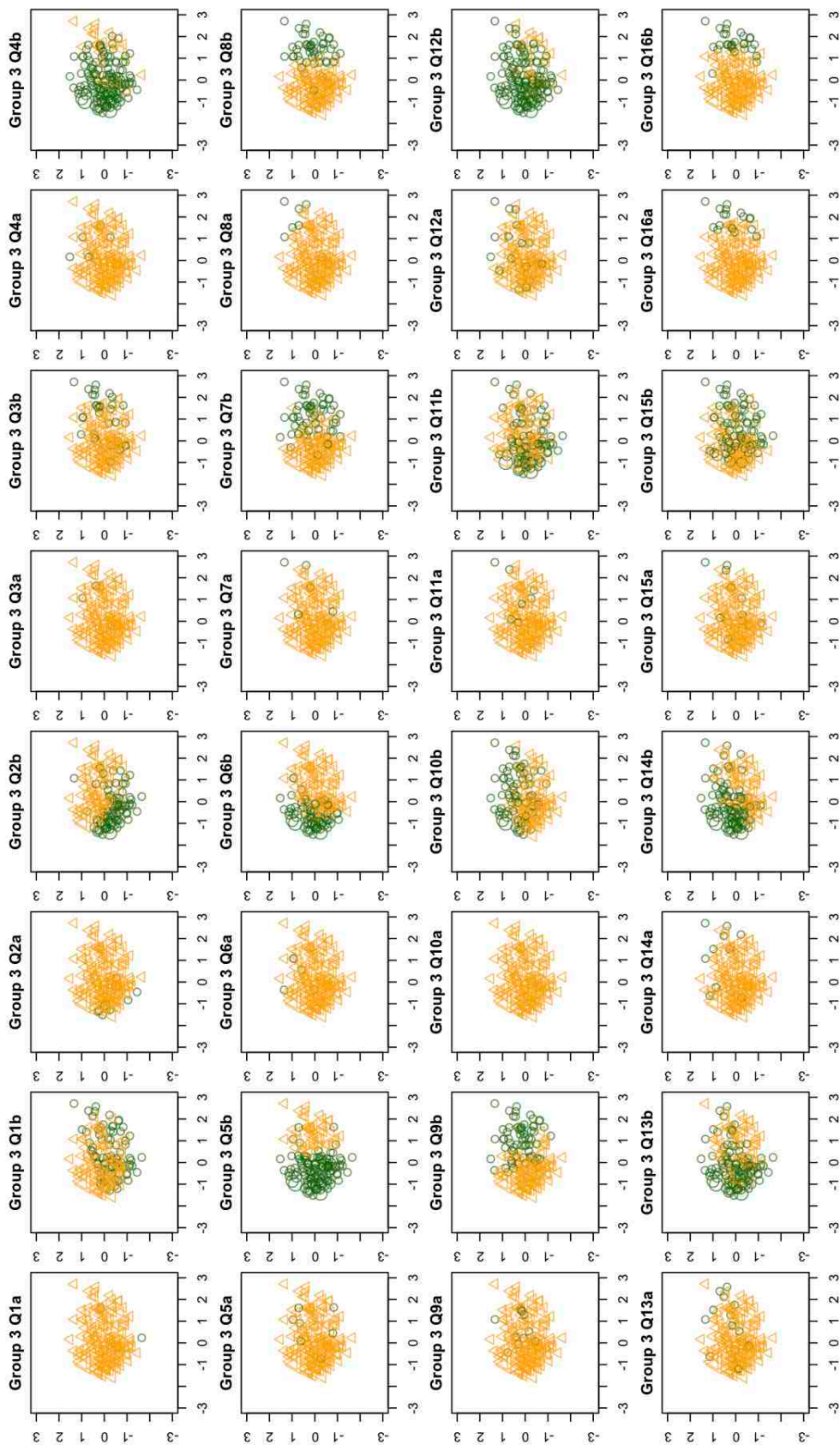
Figure 3: Plot of the $\boldsymbol{\mu}_{n3}$ for the selected model ($G = 4, D = 2, \mathbf{w}fixed$) in the U.S. congressional voting example.
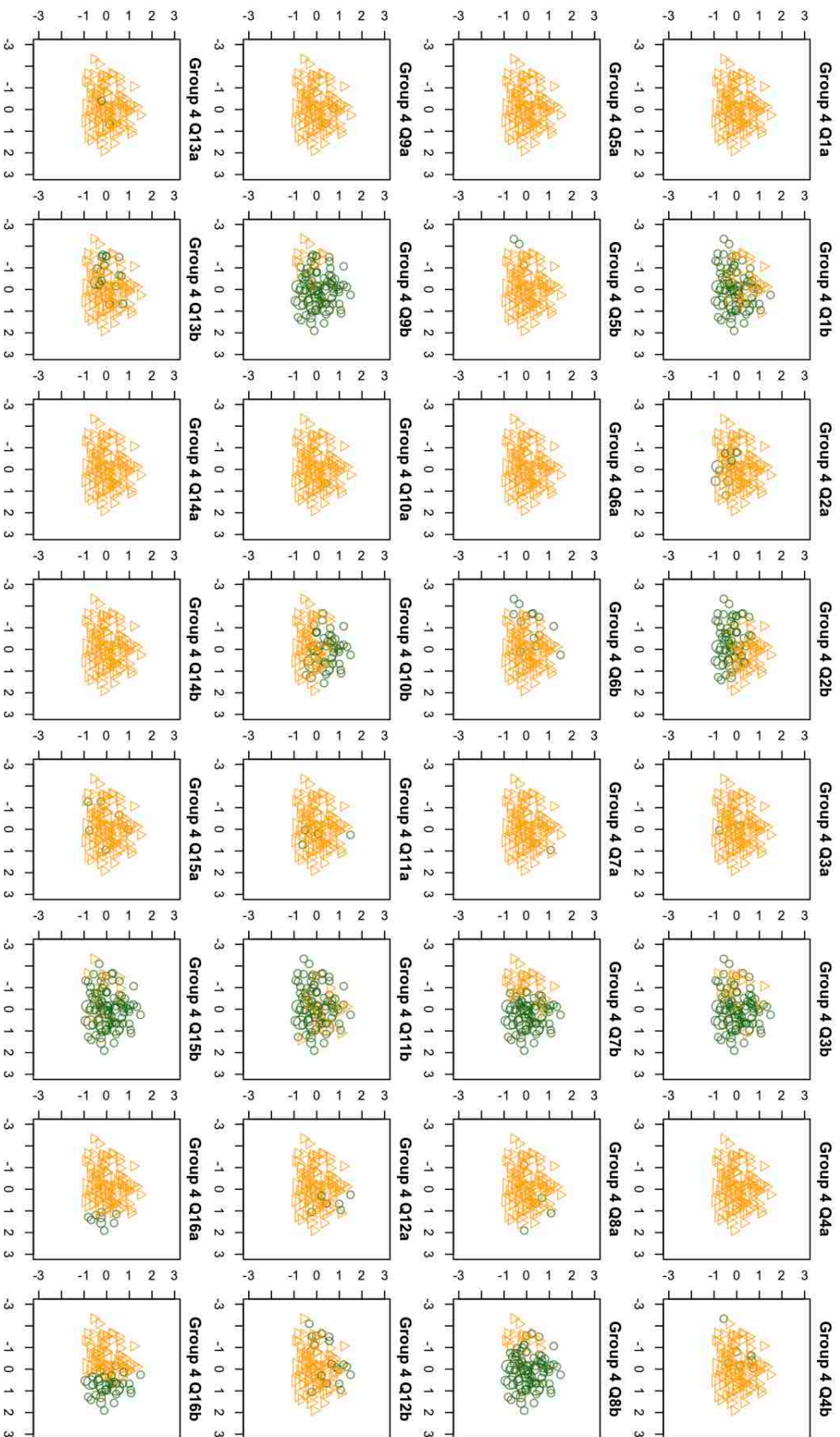
Figure 4: Plot of the $\boldsymbol{\mu}_{n_4}$ for the selected model ($G = 4$, $D = 2$, $\mathbf{w}$ *fixed*) in the U.S. congressional voting example.

8

Table 2: Estimate of the *lift* for $g = 1$ in the selected model

m	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b	9a	9b	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b	15a	15b	16a	16b
1a	1	1	1	1	1	1	1	1	0.99	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.01
1b		1	0.99	1	1.01	1	0.83	1	0.80	1	0.87	1.01	1.02	1	1	1.01	1.04	1	0.96	1	0.98	1	0.85	1	0.87	1	0.85	1	1.04	1.19	1.20
2a			1.04	1	1	1	1.06	1	1.07	1	1.04	1	0.99	1	1	1	0.98	1	0.97	1	1.02	1	1.03	1	1.04	1	1.02	1	0.98	0.89	0.88
2b				1	0.98	1	1.41	0.99	1.56	1	1.24	0.98	0.93	1	1	0.97	0.85	0.99	0.83	0.99	1.13	0.99	1.17	0.98	1.28	0.99	1.14	0.99	0.89	0.41	0.36
3a					1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3b						1	0.90	1	0.87	1	0.95	1	1.01	1	1	1	1.02	1	1	1	0.99	1	0.94	1	0.94	1	0.95	1	1.02	1.05	1.06
4a							1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4b								0.98	3.34	1	1.92	0.94	0.73	1	1	0.93	0.56	0.99	0.96	0.99	1.25	0.99	2.09	0.96	2.19	0.97	2.00	0.97	0.64	0.09	0.06
5a									0.98	1	0.99	1	1	1	1	1	1.01	1	1	1	0.99	1	0.99	1	0.99	1	0.99	1	1	1.02	1.02
5b										1	2.13	0.92	0.65	1	1	0.91	0.44	0.99	0.89	0.98	1.32	0.98	2.30	0.95	2.48	0.96	2.18	0.96	0.55	0.03	0.01
6a											1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6b												0.97	0.87	1	1	0.96	0.73	0.99	1.02	0.99	1.16	0.99	1.67	0.98	1.69	0.99	1.67	0.98	0.75	0.11	0.06
7a													1.01	1	1	1.01	1.02	1	1	1	0.99	1	0.97	1	0.96	1	0.97	1	1.01	1.04	1.04
7b														1	1	1.01	1.07	1	1.01	1	0.96	1	0.85	1.01	0.83	1	0.86	1.01	1.05	1.12	1.12
8a															1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8b																1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9a																	1.02	1	1.01	1	0.99	1	0.96	1	0.95	1	0.97	1	1.02	1.07	1.07
9b																		1	1.04	1	0.92	1	0.74	1.01	0.70	1.01	0.76	1.01	1.11	1.34	1.35
10a																			1	1	1	1	1	1	1	1	1	1	1	1.01	1.01
10b																				1	0.94	1	1.09	1.01	1	1	1.11	1.01	1.01	1.12	1.12
11a																					1	1	0.99	1	0.99	1	0.99	1	1	1.02	1.02
11b																						1	1.13	0.99	1.17	0.99	1.12	0.99	0.93	0.64	0.61
12a																							0.99	1	0.99	1	0.99	1	1	1.01	1.02
12b																								0.98	1.76	0.99	1.77	0.98	0.76	0.22	0.17
13a																									0.97	0.98	0.99	1	1.01	1.04	1.04
13b																										0.98	1.73	0.98	0.74	0.18	0.13
14a																											0.98	1	1.01	1.02	1.02
14b																												0.98	0.77	0.20	0.15
15a																													1.01	1.04	1.04
15b																														1.38	1.41
16a																															8.99

Table 3: Estimate of the *lift* for $g = 2$ in the selected model

m	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b	9a	9b	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b	15a	15b	16a	16b
1a	1.12	0.98	0.93	1.04	1.11	1.04	0.80	1.02	0.73	0.99	0.82	1.04	1.09	1.03	1.07	1.02	1.06	1.01	0.97	1.04	0.96	1.02	0.86	1.03	0.88	1.03	0.86	1.03	1.14	1.21	1.24
1b		0.96	0.89	1.10	1.30	1.09	0.53	1.06	0.41	0.96	0.53	1.09	1.12	1.08	1.17	1.06	1.15	1.03	0.86	1.09	1.03	1.06	0.68	1.06	1.07	1.06	1.13	1.05	1.04	1.11	1.11
2a			1.55	0.89	0.83	0.92	1.37	0.94	1.64	0.97	1.25	0.90	0.83	0.92	0.86	0.59	0.87	0.95	0.85	0.89	1.17	0.90	1.12	0.93	1.18	0.91	1.10	0.92	0.80	0.67	0.64
2b				0.84	0.94	1.10	0.53	1.09	3.58	1	1.76	0.71	0.53	0.74	0.58	0.59	0.61	0.87	0.61	0.71	1.53	0.73	1.34	0.80	1.53	0.74	1.28	0.79	0.51	0.29	0.25
3a					0.56	0.78	2.26	1.05	0.84	1.09	0.44	1.12	1.32	1.12	1.24	1.08	1.23	1.06	0.95	1.07	0.79	1.05	0.61	1.11	0.79	1.05	0.61	1.11	1.28	1.80	1.89
3b						1.05	0.76	1.03	0.66	1	0.80	1.06	1.11	1.04	1.08	1.03	1.08	1.02	1.01	1.06	0.93	1.04	0.80	0.87	1.04	0.86	0.87	1.04	1.16	1.25	1.28
4a							0.39	0.66	2.51	0.99	2.51	1.17	1.04	1.24	1.08	1.03	1.06	1.04	1.04	1.17	0.90	1.03	0.62	1.05	1.11	1.04	0.86	1.16	1.25	1.28	1.30
4b								0.20	5.17	0.44	0.78	0.44	0.37	0.70	0.43	0.83	0.48	0.88	0.95	0.70	1.48	0.79	2.05	0.91	1.95	0.74	1.97	0.78	0.34	0.15	0.11
5a									0.59	1.02	3.29	1.05	1.08	1.04	1.07	1.02	1.06	1.02	1.03	1.05	1.11	1.11	0.61	1.03	0.86	1.05	1.11	1.04	1.17	1.19	1.21
5b										0.73	0.86	0.78	0.63	0.43	0.83	1.03	1.09	1.03	1.03	1.48	0.94	1.04	0.91	1.03	0.90	1.04	0.91	1.03	1.12	1.19	1.21
6a											1.03	1	1	1.04	1.10	1.03	1.06	1.01	1.05	1.04	1.04	1.04	0.79	0.85	1.04	0.85	1.13	1.11	1.49	1.80	1.89
6b												0.70	1	1	1	1	1	0.82	0.84	0.57	1.80	0.67	2.55	0.66	2.43	0.59	2.40	0.67	0.17	0.03	0.02
7a													1.16	1.07	1.05	1.04	1.12	1.05	1.06	1.08	0.98	1.01	1.04	1.05	1.10	1.05	1.12	1.08	1.05	1.11	1.11
7b														1.12	1.13	1.04	1.04	1.02	1.03	0.88	0.81	1.06	1.84	0.83	1.76	0.80	1.83	0.83	0.37	0.14	0.10
8a															1.10	1.06	1.19	1.04	1.04	1.08	0.83	1.05	0.70	0.85	0.64	0.80	0.71	0.87	0.34	0.15	0.11
8b																1.06	1.10	1.02	1.03	1.11	0.91	1.08	0.81	1.05	1.06	1.07	0.81	1.08	1.26	1.38	1.41
9a																	1.06	1.04	1.08	0.84	0.98	1.03	1.01	1.04	1.01	1.04	1.04	1.03	1.09	1.15	1.16
9b																		1.04	1.11	1.11	1.01	1.08	0.84	1.03	0.90	1.09	0.91	1.03	1.03	1.12	1.13
10a																			1.04	0.84	0.76	0.74	1.34	0.73	1.05	0.75	1.06	0.75	1.26	1.38	1.41
10b																				1.08	1.01	1.08	0.81	1.08	1.05	1.09	1.10	1.08	1.27	1.38	1.41
11a																					0.88	1.07	0.85	1.05	0.83	1.07	0.86	1.05	1.22	1.37	1.41
11b																						0.90	1.20	0.92	1.24	0.90	1.18	0.92	0.75	0.61	0.58
12a																							0.92	1.04	0.89	1.05	0.93	1.05	0.75	1.24	1.24
12b																								0.89	1.52	0.90	1.62	0.89	0.54	0.35	0.30
13a																									0.88	1.04	0.90	1.04	1.13	1.21	1.23
13b																										0.86	1.04	0.88	0.55	0.34	0.29
14a																											0.89	1.05	1.15	1.24	1.26
14b																												0.90	0.53	0.34	0.28
15a																													1.14	1.24	2.17
15b																														1.26	2.33
16a																															3.48
16b																															

10

Table 4: Estimate of the *lift* for $g = 3$ in the selected model

m	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b	9a	9b	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b	15a	15b	16a	16b
1a	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1b		1	0.98	1	1.04	1	0.78	1	0.78	1	0.86	1	1.08	1.01	1.07	1	1.08	1	0.96	1	0.97	1	0.78	1.01	0.85	1	0.83	1	1.09	1.02	1.04
2a			1.01	1	1	1	1.01	1	1.02	1	1.01	1	0.99	1	0.99	1	0.99	1	0.99	1	1	1	1.01	1	1.01	1	1	1	0.99	1	1
2b				1	0.95	0.99	1.30	1	1.36	1	1.15	0.99	0.86	0.99	0.88	0.99	0.84	1	0.88	0.99	1.10	0.98	1.14	0.98	1.18	0.98	1.09	0.99	0.88	0.97	0.94
3a					1	1	0.99	1	0.99	1	0.99	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3b						1	0.81	1	0.81	1	0.91	1	1.07	1.01	1.06	1	1.07	1	1.01	1	0.97	1.01	0.86	1.01	0.89	1.01	0.91	1	1.06	1.02	1.04
4a							0.98	1	0.98	1	0.99	1	1.01	1	1.01	1	1.01	1	1	1	1	1	0.99	1	0.99	1	0.99	0.98	1.01	1	1
4b								0.99	2.16	1	1.50	0.97	0.57	0.96	0.61	0.97	0.57	1	0.97	0.99	1.18	0.97	1.88	0.96	1.65	0.96	1.58	0.98	0.63	0.85	0.74
5a									0.99	1	0.99	1	1	1	1	1	1	1	1	1	1	1	0.99	1	0.99	1	0.99	1	1	1	1
5b										1	1.56	0.97	0.53	0.97	0.58	0.97	0.52	1	0.94	0.98	1.21	0.96	1.84	0.96	1.67	0.96	1.58	0.98	0.60	0.87	0.75
6a											1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6b												0.99	0.78	0.99	0.82	0.99	0.78	1	1.01	0.99	1.10	0.99	1.42	0.98	1.35	0.98	1.35	0.99	0.78	0.95	0.91
7a													1.01	1.01	1.01	1	1.01	1	1.01	1	1.01	1	0.98	1	0.99	1	0.99	1	1.01	1	1.01
7b														1.01	1.16	1.01	1.19	1	1.03	1.01	0.92	1.01	0.69	1.02	0.74	1.01	0.78	1.01	1.16	1.05	1.09
8a															1.01	1.01	1.01	1	1	1	0.99	1	0.98	1	0.98	1	0.99	1	1.01	1	1.01
8b																1.01	1.16	1	1.02	1	0.93	1.01	0.72	1.01	0.78	1.01	0.81	1.01	1.13	1.05	1.08
9a																	1.01	1	1	1	0.99	1	0.98	1	0.98	1	0.99	1	1.01	1	1.01
9b																		1	1.04	1.01	0.91	1.02	0.70	1.02	0.74	1.02	0.78	1.01	1.16	1.05	1.09
10a																			1	1	1	1	1	1	1	1	1	1	1	1	1
10b																				1	0.96	1.01	1.07	1.01	1	1.01	1.06	1.01	1.01	1.01	1.01
11a																					1	1	0.99	1	0.99	1	0.99	1	1.01	1	1
11b																						0.99	1.11	0.99	1.11	0.99	1.08	1	0.93	0.98	0.96
12a																							0.98	1	0.98	1	0.99	1	1.01	1	1.01
12b																								0.97	1.51	0.98	1.53	0.99	0.71	0.90	0.82
13a																									0.98	1	0.98	0.99	1.01	1	1.01
13b																										0.98	1.37	0.99	0.76	0.93	0.87
14a																											0.99	1	1.01	1	1.01
14b																												0.99	0.78	0.94	0.89
15a																													1.01	1	1
15b																														1.04	1.07
16a																															1.03

Table 5: Estimate of the *lift* for $g = 4$ in the selected model

m	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b	9a	9b	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b	15a	15b	16a	16b
1a	1																														
1b		1	0.96	1	1.43	1	0.97	1	0.98	0.91	1.01	1	1.45	1.01	1.62	1	1.50	1	0.92	1	0.92	1	0.93	1	0.93	1	1	1	1.42	1.05	1.21
2a			1.02	1	0.97	1	1.02	1	1	1	1	1	0.96	1	0.94	1	0.95	1	0.99	1	1.02	1	1	1	1	1	1	1	0.97	0.97	0.98
2b				1	0.67	1	1	1	1.01	1.04	1.01	1.01	0.59	0.98	0.47	1	0.48	1	0.87	1	1.20	1	1.03	1	1.04	1	1	1	0.69	0.95	0.79
3a					1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3b						1	0.93	1	0.95	0.82	0.82	1.01	2.11	1.02	2.69	1	2.46	1	1.03	1	0.73	1	0.88	1	0.86	1	1	1	1.97	1.08	1.40
4a							1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4b								1	1.01	1.01	1.01	1	0.92	1	0.87	1	0.89	1	1	1	1.02	1	1.01	1	1.01	1	1	1	0.93	0.97	0.97
5a									1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5b										1.01	1.01	1	0.95	1	0.90	1	0.91	1	1	1	1.01	1	1.01	1	1.01	1	1	1	0.95	1	0.99
6a											1	1	0.81	1	0.69	1	0.74	1	1	1	0.61	1	0.84	1	0.80	1	1	1	0.82	0.94	0.94
6b												1.01	1	1	1.01	1	1.01	1	1.12	1	1.01	1	1.01	1	1.01	1	1	1	2.43	1.09	1.51
7a													1.02	1.02	2.85	1.01	2.63	1	1.07	1	0.67	1	0.87	1	0.85	1	1	1	2.08	1.09	1.48
7b														1.03	1.03	1	1.03	1	1.01	1	0.99	1	1	1	1	1	1	1	1.01	1.01	1.01
8a															1.01	1.01	3.65	1	1.10	1	0.59	1	0.81	1	0.77	1	1	1	2.66	1.09	1.55
8b																1	1.01	1	1.01	1	1.01	1	1.01	1	1.01	1	1	1	1.02	1.01	1.02
9a																	1.01	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9b																		1	1.12	1	0.61	1	0.84	1	0.80	1	1	1	2.43	1.09	1.51
10a																			1	1	1	1	1	1	1	1	1	1	1	1	1
10b																				1	0.93	1	1.01	1	1	1	1	1	1.03	1.03	1.03
11a																					1	1	1	1	1	1	1	1	1.01	1.01	1.01
11b																						1	1.03	1	1.04	1	1	1	0.74	0.95	0.82
12a																							1	1	1	1	1	1	1	1	1
12b																								1	1.02	1	1	1	0.88	0.99	0.95
13a																									1	1	1	1	1.01	1	1
13b																										1	1	1	0.86	0.99	0.94
14a																											1	1	1	1	1
14b																												1	1.02	1	1
15a																													1	1	1
15b																														1.01	1.39
16a																															1.07
16b																															