LATENT VARIABLE MODELS FOR MIXED MANIFEST VARIABLES

Irini Moustaki

London School of Economics and Political Science

University of London

Submitted in Fulfilment of the Requirement for the Degree of Doctor of Philosophy

Acknowledgements

I am most grateful to my supervisor Colm O'Muircheartaigh who provided me with continuous support and encouragement.

I am indebted to Martin Knott for his support and interest shown during my studies.

The Statistics Department and the Methodology Institute at the LSE provided a friendly and supportive environment for which I am grateful.

I thank Tetti for being a very close and encouraging friend and Dimitra for our long discussions. I have also forgiven Nancy for the continuous interruptions.

Finally, I would like to acknowledge the Onassis Public Benefit Foundation and the Economic and Social Research Council for supporting my studies financially.

Στη μητερα μου και στη θεια μου

Abstract

Latent variable models are widely used in social sciences in which interest is centred on entities such as attitudes, beliefs or abilities for which there exist no direct measuring instruments. Latent modelling tries to extract these entities, here described as latent (unobserved) variables, from measurements on related manifest (observed) variables. Methodology already exists for fitting a latent variable model to manifest data that is either categorical (latent trait and latent class analysis) or continuous (factor analysis and latent profile analysis).

In this thesis a latent trait and a latent class model are presented for analysing the relationships among a set of mixed manifest variables using one or more latent variables. The set of manifest variables contains metric (continuous or discrete) and binary items. The latent dimension is continuous for the latent trait model and discrete for the latent class model.

Scoring methods for allocating individuals on the identified latent dimensions based on their responses to the mixed manifest variables are discussed.

Item nonresponse is also discussed in attitude scales with a mixture of binary and metric variables using the latent trait model.

The estimation and the scoring methods for the latent trait model have been generalized for conditional distributions of the observed variables given the vector of latent variables other than the normal and the Bernoulli in the exponential family.

To illustrate the use of the mixed model four data sets have been analyzed. Two of the data sets contain five memory questions, the first on Thatcher's resignation and the second on the Hillsborough football disaster; these five questions were included in BMRBI's August 1993 face to face omnibus survey. The third and the fourth data sets are from the 1990 and 1991 British Social Attitudes surveys; the questions which have been analyzed are from the sexual attitudes sections and the environment section respectively.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis we develop a method for analysing latent variable models with binary and metric manifest variables when observations may be missing. Binary items can take only two possible values (agree / disagree) while metric variables have real number values and can be either discrete or continuous.

Latent variable models are widely used in social sciences in which interest is centred on entities such as attitudes, beliefs or abilities for which there exist no practical direct measuring instruments. Latent modelling tries to extract these entities, here described as latent (unobserved) variables, from measurements on related manifest (observed) variables. These latent variables are entities that in practice we may not be able or willing to directly measure e.g. wealth, status or which are not directly measurable such as attitude and ability.

Bartholomew (1987) presented a unified approach for treating latent variable models. As already mentioned, in the theory of latent variable models we distinguish between two types of variables, the observed or manifest variables and the unobserved or latent variables. Both types of variables can be either metric or categorical. When both the manifest and latent variables are metric we apply factor analysis, when both are categorical we apply latent class analysis, metric manifest and categorical latent gives latent profile analysis and finally for categorical manifest and metric latent we apply latent trait analysis.

In the literature there are two approaches for estimating the parameters of these kinds of models. First there is the underlying variable approach which treats all manifest variables as continuous by assuming that underlying each categorical manifest variable there is a continuous unobserved variable. Secondly there is the response function approach which starts by defining for each individual in the sample the probability of responding positively to a variable given the individual's position on the latent factor space.

We will concentrate on the case where the manifest variables are of mixed type (binary and metric) and the latent variables are either continuous or discrete. In the literature only the underlying variable approach has been used for estimating the 'mixed' model. We will develop a response function approach for the 'mixed' model but before that a review of the existing approaches for binary, metric and mixed manifest variables will be given.

1.1 Notation

Variables that are directly observed are known as *manifest variables* and variables that are unobserved are known as *latent variables.* The manifest variables will be denoted by x and the vector $\mathbf{x}$ of dimension $(p \times 1)$ will denote a group of manifest variables. The vector $\mathbf{x}$ can contain both metric and binary items. The metric manifest variables will be denoted by w and the vector $\mathbf{w}$ of dimension $(r \times 1)$ will denote a group of metric manifest variables. The binary manifest variables will be denoted by v and the vector $\mathbf{v}$ of dimension $(s \times 1)$ will denote a group of binary manifest variables. In all the chapters of this thesis the above notation will be adopted for metric and binary manifest variables.

The subscription h on the vector $\mathbf{x}$, $\mathbf{v}$ or $\mathbf{w}$ will denote the values of the manifest variables for the hth sample member and the subscription i will denote the ith manifest variable.

Latent variables with standard normal distributions will be denoted by z and y otherwise. The number of such variables will be q.

1.2 Factor analysis for metric manifest variables

The origins of factor analysis go back to Spearman (1904). He tried to see whether something like 'general intelligence' could explain the correlations among sets of test scores. For an overview of the origins and the development of factor analysis see Bartholomew (1995) The first statistical treatment of factor analysis was given in Lawley and Maxwell (1971).

In factor analysis we have a number r of observed metric variables that we want to express as linear combinations of q latent variables where q is much less than r. In other words the object of the analysis is to explain the interrelationships among a number of r manifest variables using a number of latent variables q where $q < r$. This analysis is carried out through the covariance or the correlation matrix of the r manifest variables w.

Suppose $w_1, \cdots, w_r$ are r metric variables. The linear factor model is written

$$w_i = \mu_i + \sum_{j=1}^{q} \lambda_{ij} z_j + e_i \quad i = 1, \cdots, r \tag{1.1}$$

or in a matrix form

$$\mathbf{w} = \mu + \mathbf{\Lambda z} + \mathbf{e}$$

where $\mathbf{e} \sim N_r(\mathbf{0}, \mathbf{\Psi})$.

Suppose that the latent variables follow independent standard normal distributions, $\mathbf{z} \sim N_q(\mathbf{0}, \mathbf{I})$. Under the assumption of conditional independence, which states that conditional on the vector of latent variables $\mathbf{z}$ the responses to the r manifest items are independent, the conditional distribution of $\mathbf{w}$ given $\mathbf{z}$ follows a normal distribution $N_r(\mu + \mathbf{\Lambda z}, \mathbf{\Psi})$, where $\mathbf{\Lambda}$ is $r \times q$ matrix of coefficients (factor loadings) and $\mathbf{\Psi}$ is a $r \times r$ diagonal matrix of specific variances. Considering the above results the conditional distribution of each manifest variable w_i given $\mathbf{z}$ becomes:

$$g(w_i \mid \mathbf{z}) = (2\pi)^{-1/2} \Psi_{ii}^{-1/2} \exp(-\frac{1}{2\Psi_{ii}}(w_i - \mu_i - \sum_{j=1}^{q} \lambda_{ij} z_j)^2), \quad q \geq 1$$

It then follows that the manifest variables $\mathbf{w}$ follow a normal distribution $N_r(\mu, \mathbf{\Lambda\Lambda'} + \mathbf{\Psi})$. The posterior distribution of $\mathbf{z}$ given $\mathbf{w}$ follows a normal distribution, (see Basilevsky 1994, chapter 6):

$$N_q(\mathbf{\Lambda}'\mathbf{\Sigma}^{-1}(\mathbf{w} - \mu), (\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda} + \mathbf{I})^{-1})$$

where

$$\mathbf{\Sigma} = \mathbf{\Lambda\Lambda}' + \mathbf{\Psi}$$

and $q < r$. The matrix $\mathbf{\Lambda}_{r \times q}$ contains the covariances between elements of $\mathbf{z}$ and $\mathbf{w}$.

1.2.1 Estimation Methods

The estimation of the model parameters is based on the maximization of the loglikelihood of the marginal distribution of the manifest variables. Details of the estimation method are given in Lawley and Maxwell (1971) and Bartholomew (1987).

Rubin and Thayer (1982) and (1983) proposed an EM algorithm for finding maximum likelihood estimates for the parameters of the linear factor model. The EM algorithm was introduced by Dempster, Laird, and Rubin (1977) for maximum likelihood estimation in a multivariate model with missing data. The EM algorithm for our application treats the latent variables $\mathbf{z}$ as missing data, and iteratively maximizes the likelihood supposing $\mathbf{z}$ were observed.

Other estimation methods in the literature are the unweighted least squares method (ULS) and the generalized least squares method (GLS). Both methods try to fit a model by choosing estimates so that the observed and theoretical correlation matrices are as close as possible.

With the ULS one minimizes the function:

$$(S - \hat{\Sigma})'(S - \hat{\Sigma})$$

and the GLS minimizes the function:

$$(S - \hat{\Sigma})'W^{-1}(S - \hat{\Sigma})$$

where S and $\hat{\Sigma}$ are the sample and estimated by the model covariance matrix respectively and W is a weight matrix. These methods will be discussed later for factor analysis with dichotomous variables.

All three estimation methods, ML, ULS and GLS, provide consistent estimates of the parameters $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ and large sample chi-square tests for the goodness-of-fit.

Unique estimates of the parameters $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ do not exist if we do not impose some constraints on the parameters for the case where $q > 1$. One possible constraint is to take $\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ diagonal. This is as a mathematical convenience but also it makes the conditional distribution of the z's given w independent.

Any orthogonal rotation of the factors in the q-space will give a new set of factors which will also satisfy $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$. Thus the likelihood will have more than one maximizing value and once one is found others can be found by orthogonal transformation of the solution obtained. This is what is called rotation in factor analysis. Rotation of the factor solution allows for an easier interpretation of the factors although in cases where there is a very dominant general factor it may not be helpful.

Asymptotic standard errors for the parameters can be obtained by inverting the expected information matrix at the maximum likelihood solution. Lawley and Maxwell (1971) give the variance covariance matrices for the estimated parameters. These formulae assume that the estimates have been obtained from the covariance matrix rather than the correlation matrix and so they are not applicable in practice. Resampling techniques such as jackknife and bootstrapping can be used for calculating standard errors.

1.2.2 Goodness of fit

A test for the fit of the linear factor model may be based on the likelihood ratio statistic. If the number of latent variables q has been specified a priori, we use the likelihood ratio statistic to test the null hypothesis that $\mathbf{\Sigma}$ is $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$ against the alternative that $\mathbf{\Sigma}$ is unconstrained.

It is known that for large samples the statistic $-2\{L(H_0) - L(H_1)\}$ under the null hypothesis is distributed approximately as χ^2 with degrees of freedom

$$\tfrac{1}{2}r(r+1) - (rq + r - \tfrac{1}{2}q(q-1)) = \tfrac{1}{2}\{(r-q)^2 - (r+q)\}$$

The number of degrees of freedom is the number of parameters in $\mathbf{\Sigma}$ minus the number of linear constraints imposed by the null hypothesis.

The loglikelihood under the null hypothesis is

$$L(H_0) = -\frac{1}{2}nr\log 2\pi - \frac{1}{2}n\log \mid \hat{\mathbf{\Sigma}} \mid -\frac{1}{2}n\mathrm{tr}\hat{\mathbf{\Sigma}}^{-1}\mathbf{S}$$

The loglikelihood under the alternative hypothesis is

$$L(H_1) = -\frac{1}{2}nr\log 2\pi - \frac{1}{2}n\log \mid \mathbf{S} \mid -\frac{1}{2}n\mathrm{tr}\mathbf{S}^{-1}\mathbf{S}$$

The statistic is then

$$-2\{L(H_0) - L(H_1)\} = n\{\mathrm{tr}\hat{\mathbf{\Sigma}}^{-1}\mathbf{S} - \log \mid \hat{\mathbf{\Sigma}}^{-1}\mathbf{S} \mid - \mathrm{r}\} \tag{1.2}$$

Where the $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}}$

Bartlett (1954) showed that the likelihood ratio statistic can be better approximated from the chi-square distribution by replacing n in equation (1.2) by $n - \frac{2r+11}{6} - \frac{2q}{3}$

If q has not been specified in advance then a procedure for choosing the best value for q will be to start with $q = 1$ and stop when the likelihood ratio statistic is not significant. Lawley and Maxwell (1971) mentioned that the above procedure for choosing q does not take into account the fact that a sequence of hypotheses is being tested, with each one dependent on the rejection of all predecessors and the significance level has not been adjusted. Other criteria for choosing the value of q are mentioned in Bartholomew (1987).

1.3 Factor analysis for binary manifest variables

In the literature of latent variable models there are two approaches for the fit of latent models on binary items. One is called the underlying variable approach and it is an extension to the theory described above for continuous variables and the other one is called the response function approach. The first method supposes that the binary manifest variables have been produced by dichotomizing underlying continuous variables. The second method defines a response function that gives the probability of a positive response for an individual with latent position z. Bartholomew (1987) showed that the underlying variable approach and the response function approach are equivalent for binary items but different for polytomous items.

1.3.1 Underlying variable approach

This approach brings the analysis of binary variables within the framework of factor analysis for metric manifest variables. This is achieved by assuming that each binary variable is generated by an underlying continuous variable in the following way:

$$v_i = \begin{cases} 1 & \text{if } v_i^* \geq \tau_i \\ 0 & \text{if } v_i^* < \tau_i \end{cases}$$

where τ_i are called threshold parameters and

$$\mathbf{v}^* = \mu + \mathbf{\Lambda z} + \mathbf{e}$$

Hence, if we make the same assumptions as before for the distributions of $\mathbf{z}$ and $\mathbf{e}$ then the linear factor model can be fitted on the covariance or correlation matrix of the $\mathbf{v}^*$'s variables. There are maximum likelihood methods for estimating the correlation coefficients from a 2x2 cross-classification of the data. These are called tetrachoric correlation coefficients.

The assumption of normality for the underlying response variables might not be appropriate all the time and with all the variables. There are items for which there are no direct meaningful underlying variables to consider. This set of assumptions just makes the analysis of binary response consistent with factor analysis for continuous variables.

Other types of distribution for the bivariate continuous distribution underlying each of the 2x2 observed tables can be considered such as the C-type distribution. For this distribution, whatever the threshold is, the cross product ratio is the same. This property is quite important especially when the threshold values have been defined arbitrarily, and so the results from the analysis do not depend on these cut values.

Christofferson (1975) fitted a linear factor model on a set of s binary variables by estimating the parameters which minimize the distance between the observed and expected first- and second-order marginal proportions assuming that the underlying variables $\mathbf{v}^*$ follow a multivariate normal distribution.

Let P_i and P_{ij} be the expected proportion who respond positively to item i and positively to items i and j respectively. His method is based on minimizing the differences between these expected proportions and the ones observed from the sample denoted by lower letters p_i and p_{ij}.

The expected proportions are defined as:

$$P_i = \int_{-\infty}^{\tau_i} f(u)du \tag{1.3}$$

and

$$P_{ij} = \int_{-\infty}^{\tau_i} \int_{-\infty}^{\tau_j} f(u_1, u_2; \rho_{ij}) du_1 du_2 \tag{1.4}$$

where $f(u)$ and $f(u_1, u_2)$ are the standard univariate and bivariate normal density function.

The model is written as:

$$p_i = P_i + \epsilon_i, \qquad i = 1, \cdots, s$$

$$p_{ij} = P_{ij} + \epsilon_{ij}, \qquad i = 1, \cdots, s-1, \quad j = i+1, \cdots, s$$

or in a vector form:

$$\mathbf{p} = \mathbf{P} + \epsilon$$

where the error term ϵ has expectation zero and covariance matrix $\mathbf{\Sigma}_\epsilon$. The expected proportions $\mathbf{P}$ are expressed in terms of the thresholds, tetrachoric correlations, the factor loadings and the error term ϵ.

When the model is true the differences $\epsilon = \mathbf{p} - \mathbf{P}$ for the single and pair of items will follow a multivariate normal distribution in large samples with mean zero and covariance matrix $\mathbf{\Sigma}_\epsilon$. Christofferson (1975) obtained a consistent estimator of this matrix S_e, this estimator uses also information from third and fourth order marginal proportions, (Christofferson appendix 2) and proceeded to minimize the generalized least squares quantity

$$\mathbf{Q} = (\mathbf{p} - \mathbf{P})'\mathbf{S}_\mathbf{e}^{-1}(\mathbf{p} - \mathbf{P})$$

This function was minimized using the Fletcher and Powell method. The estimators obtained by this method are asymptotically efficient among those estimators that use the same amount of information, i.e. first and second order probabilities. A chi-square test for the goodness-of-fit of the model is available and standard errors of the estimates are also available from the inverse of the matrix of second derivatives of Q respect to the parameters. The generalized least square method has an advantage over the full maximum likelihood method when we fit several latent variables.

Muthén (1978) proposed a transformation of Christofferson's method which is also based on the first and second order proportions but simplifies the computations because it avoids the integrations needed for the calculation of the first and second marginal proportions, equations (1.3) and (1.4) respectively.

The estimates obtained are asymptotically efficient among those estimators that use the same amount of information. Muthén's method is less computationally heavy than Christofersson's but it is still limited to 20-25 items due to the rapid increase of the dimension of the weight matrix with the increase of the items. These methods have been implemented in a computer program LISCOMP (Muthén 1987).

Christofferson (1975) and Muthén (1978) methods presented above make the assumption that most of the information needed in the analysis is contained in the first- and second-order margins. This assumption comes from the fact that if the underlying variables, v_i^*, were known then the sample covariance matrix is required for the estimation of the model parameters and these require a knowledge of the bivariate distributions.

1.3.2 Response function approach

The methods discussed in the previous section were all 'limited information' methods in the sense that they take account of the first and second order probability margins. The only methods that use all the information provided from the 2^s response patterns are the maximum likelihood based methods and they will be discussed here. But before we go into the ML estimation method we discuss the different types of response function.

Response function

The response function is denoted by $\pi_i(\mathbf{y})$ and gives the probability that an individual will respond positively to item i given his latent position y. There are many different models for the response function. Bartholomew (1980) defines some desirable properties for the response function. First, the response function $\pi(\mathbf{y})$ must be monotonic non-increasing or non-decreasing with respect to the latent variable. Secondly, if $\pi(\mathbf{y}) \in \mathcal{F}$ then the function obtained by replacing any sub-set of the elements of $\mathbf{y}$ by their complements should also belong to $\mathcal{F}$. This property satisfies the arbitrariness of the direction in which the latent variable is measured. Thirdly, if $\pi(\mathbf{y}) \in \mathcal{F}$ then $1 - \pi(\mathbf{y}) \in \mathcal{F}$. This property satisfies the arbitrariness in the direction of the ordering of the categories. There are some other properties that deal with special cases, the complete independence, $(\pi(\mathbf{y}) = \pi \in \mathcal{F})$, and the case of a perfect scale, (Guttman). Some less formal properties are the flexibility of the response function of describing many different shapes and the small number of parameters to be estimated.

Bartholomew considered the class of functions:

$$G^{-1}(\pi_i(\mathbf{y})) = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} H^{-1}(y_j)$$

where $\pi_i(\mathbf{y})$ is the response function, the probability that an individual will respond positively to item i given latent position $\mathbf{y}$, and y_j $j = 1, \cdots, q$ have independent uniform distributions with mean zero and variance one.

The functions G^{-1} and H^{-1} are chosen so that the response function has the properties described above.

In the literature the most commonly used response functions are the logit ($G^{-1} = \text{logit}(v) = \log v/(1-v)$) and the probit ($G^{-1} = \text{probit}(v) = \Phi^{-1}(v)$), where Φ is the cumulative of the standard normal). Lord and Novick (1968) use the logit/probit, model which has the logit for G^{-1} and the probit for H^{-1}. Bock and Aitkin (1981) use the probit model in which the probit is selected for both G^{-1} and H^{-1}. Bartholomew (1980) prefers the logit/probit model, (briefly called logit) in which G^{-1} is selected to be the logit and H^{-1} is selected to be the probit because it is easy to estimate and it also gives some very useful results when it comes to the scoring methods.

In this thesis the logit response function will be used but a brief overview of the others will be given. The logistic and the normal are very similar in shape and which one is going to be used is a matter of practicality. There is an approximate relationship between the logistic and the normal which is given by:

$$\text{logit}(v) \approx \frac{\pi}{\sqrt{3}} \Phi^{-1}(v)$$

Hence the logit/probit model:

$$\text{logit}[\pi_i(\mathbf{y})] = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} \Phi^{-1}(y_j)$$

is approximately the same as:

$$\text{logit}[\pi_i(\mathbf{y})] = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} (\sqrt{3}/\pi) \text{logit} y_j$$

So if we estimate the item parameter α_{ij} of the logit/logit model we can also get approximately the item parameter α_{ij} for the logit/probit model by multiplying that by $\sqrt{3}/\pi$.

Similarly from the probit model:

$$\Phi^{-1}[\pi_i(\mathbf{y})] = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} \Phi^{-1}(y_j)$$

we can get the logit/probit model:

$$\text{logit}[\pi_i(\mathbf{y})] = \pi/\sqrt{3}[\alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} \Phi^{-1}(y_j)]$$

Finally, Bartholomew (1987) showed that in order to allow rotation of the factor solution it is appropriate to transform the variables y_j to normally distributed variables z_j, ($z_j = H^{-1}(y_j)$).

In this thesis the shape of the response function is taken to be the logit function:

$$\text{logit}\pi_i(\mathbf{z}) = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j \quad q \geq 1 \tag{1.5}$$

Interpretation of the parameters

The parameters α_{i0} and α_{ij} define the shape of the response function which shows how the probability of a correct response increases with 'ability' and so it should be monotonic nondecreasing in the latent space.

Working with the logit model, the coefficient α_{i0} is the value of the logit$\pi_i(\mathbf{z})$ at $\mathbf{z} = \mathbf{0}$. In other words this is the probability of a positive response for the median individual. The α_{i0} are called *difficulty* parameters. Items with large difficulty parameters are expected to be answered the same by most of the individuals.

The coefficient α_{ij} is a measure of the extent to which the ith manifest variable discriminates between individuals. For two individuals with different positions on the latent dimension z_j, the bigger the absolute value of α_{ij} the greater the difference in their probabilities of giving a positive response to item i and thus the easier to discriminate between them on the evidence of their responses to item i. These *discrimination* parameters play a very important role since they give a different weight to each item according to its discriminating power. We will discuss that again when we talk about the scaling methods.

Ideally, when we construct scales in order to measure a particular concept all the items should have the same discriminating power. But in reality it often turns out that the discrimination parameters take very large values for some of the items. That means that the response function has a threshold for this item. Albanese (1990) investigated the behaviour of the likelihood for the one factor logit/probit model when some of the items have large discriminating parameters. She suggested a reparametrization of these α_{ij} coefficients:

$$\alpha^*_{ij} = \alpha_{ij}/(\sqrt{\sum_{j=1}^{q} \alpha^2{}_{ij} + 1})$$

which gave useful results in the sense that it showed better behaviour of the likelihood function.

Properties of the response function

There are two properties that one should have in mind when following this approach.

1. For binary items the outcome which is going to be regarded as 'correct' or 'wrong' is totally arbitrary. So if the correct answer has probability $\pi_i(\mathrm{z})$ then the wrong answer has probability $1 - \pi_i(\mathrm{z})$, where:

$$\begin{aligned} \pi_i(\mathrm{z}) &= \frac{\exp(\alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j)}{1 + \exp(\alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j)} \\ &= \{1 + \exp(-\alpha_{i0} - \sum_{j=1}^{q} \alpha_{ij} z_j)\}^{-1} \end{aligned}$$

and

$$1 - \pi_i(\mathrm{z}) = \{1 + \exp(\alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j)\}^{-1}$$

These two equations mean that if we increase z then the probability of a correct response increases by the same amount that the probability of a wrong response decreases.

2. The direction of the latent variables is arbitrary. For the logit model a simple

rotation of factors changing the z_j by $-z_j$ has results in a change of the signs of the α_{ij} parameters.

Estimation methods

There are three different maximum likelihood methods in the literature. These are the methods of joint maximum likelihood, conditional maximum likelihood (CML) and marginal maximum likelihood (MML).

The joint and the conditional maximum likelihood are called "fixed effects" solutions, they assume that abilities are fixed parameters and are finite in number, when in fact they are not identifiable and have a distribution over the population of subjects. So each individual's position on the latent scale is represented by a parameter.

The MML is a random effects solution in which individuals are supposed to be sampled at random from some population and so each individuals' position on the latent scale is the value of a random variable.

In the fixed effects model the number of parameters to be estimated is much more than in the random effects model and as the sample size increases the number of parameters increases proportionally. Estimation becomes a difficult task and the random effects model seems preferable.

The joint and the conditional ML methods have been used for estimating the parameters of the Rasch model. The Rasch model can be obtained from the two parameter logistic model, (see equation 1.5), when $q = 1$ and $\alpha_{i1} = \alpha$ by treating also each individual's position on the latent scale as a fixed parameter instead of a random variable.

The joint maximum likelihood is based on the simultaneous estimation of the item parameters and the person abilities. Haberman (1977) has shown that consistent estimates of the Rasch difficulty parameters are obtained by the joint method as both the number of items and the number of subjects increases without limit, but this condition is not realistic in practice. Several researchers have avoided this problem by assuming that subjects who have the same number of right score or the same score pattern or who have been assigned provisionally to homogeneous groups have the same ability.

The conditional maximum likelihood has been used for estimating the difficulty parameters of the Rasch model. The CML method is based on the conditional distribution given minimal sufficient statistics. For example for the Rasch model instead of maximizing the likelihood function with respect to the difficulty parameters and the individual's ability parameters, with the CML method we maximize the conditional likelihood of the item parameters given minimal sufficient statistics for the individual's ability parameters which in that model are the raw score for each individual. Andersen (1970) and (1972) give the asymptotic properties of CML and the CML estimators for the one parameter logistic model (Rasch model). The method cannot be applied to the two parameter logistic model because the sufficient statistics depends on the discrimination parameters. Hence if the discrimination parameters are not estimated with a reliable way the CML cannot be used any more.

The MML approach has been discussed more than the other two approaches.

Bock and Lieberman (1970) fitted a response model on a number of binary items using an unconditional maximum likelihood estimation of a two parameter probit model on the assumption that individuals are a random sample from a standard normal distribution of ability.

Their approach is like assuming again that underlying each manifest variable there is an underlying continuous variable. But by estimating the threshold and item parameters with ML it avoids the case of estimating a tetrachoric correlation matrix which might not be positive definite.

The maximum likelihood solution was obtained via a Newton-Raphson method and Gauss-Hermite integration. Their method had computational difficulties due to the computations required in the Newton-Raphson method which limited the number of items to be analyzed to 10 or 12 and the number of factors to one.

Bock and Aitkin (1981) reformulated the Bock and Lieberman likelihood equations to make the estimation method more computational attractive. Their approach is based on a variation of the EM algorithm. In this formulation of the problem the distribution of the latent variable does not need to be known in advance instead it can be estimated as a discrete distribution on a finite number of points. By defining each individual's position on the latent dimension the item parameters can be estimated using probit analysis.

Their method applies to more than one latent dimension and it provides full-information factor analysis of dichotomous and polytomous items. Because it uses the probit as response function it lacks the sufficiency principle which is described in Bartholomew (1987).

Bartholomew (1987) uses the same formulation of the estimation procedure, taking as response function the logit instead of the probit. Since in this thesis the logit response function will be taken for the binary items, Bartholomew's formulation will be described in more details.

Marginal maximum likelihood

Suppose $v_1, v_2, \cdots, v_s$ are s binary items taking values 0 and 1. Let v_{ih} be the value of the h^{th} individual for the i^{th} item, $(h = 1, \cdots, n)$. The row vector $v_h' = (v_{1h}, \cdots, v_{sh})$ is referred to as the response pattern of the h^{th} individual.

First the results of the one factor model will be presented. If we would think of estimating parameters for any model which takes account of the manifest variables v and the latent variable z it would be appropriate to start with the distribution of the manifest variables $\mathbf{v}$ because that is the one we observe.

$$f(\mathbf{v}) = \int_{-\infty}^{+\infty} g(\mathbf{v} \mid z) h(z) dz$$

Hence the only concern now is to define the form of the conditional distribution $g(\mathbf{v} \mid z)$ and to make an assumption about the distribution of the latent variables.

Under the assumption of conditional independence

$$g(\mathbf{v} \mid z) = \prod_{i=1}^{s} g_i(v_i \mid z)$$

This assumption means that the set of latent variables is complete and so it explains perfectly the interrelationships among the s manifest items. This is often called the assumption of *conditional* or *local independence*. This is an assumption that cannot be tested empirically because there is no way to keep z constant.

Since v_i's are binary

$$g(v_i \mid z) = \pi_i(z)^{v_i}(1-\pi_i(z))^{1-v_i} \quad i = 1, \cdots, s$$

where $\pi_i(z) = Pr(v_i = 1 \mid z)$ is the probability of a positive response for an individual with latent position z and is called response function. The mathematical properties of the response function discussed in a previous section.

The question is what form the conditional distribution must take in order for the reduction of the dimensionality from s to q to be possible, or in other words the posterior distribution of z given $\mathbf{v}$ to depend on $\mathbf{v}$ only through a q function of $\mathbf{v}$. This reduction is named sufficiency principle by Bartholomew (1987) and it will explained in more details in the next chapter.

Barankin and Maitra (1963) give the necessary and sufficient conditions, which are required in order that reduction of the dimensionality of the data to be possible. These conditions require that at least $(s-q)$ of the g's must have the exponential form. By choosing the response function to be the logit the above requirement is satisfied. The response function takes the form:

$$\text{logit}\pi_i(z) = \alpha_{i0} + \alpha_{i1}z \tag{1.6}$$

The latent variable is assumed to have a standard normal distribution. For a random sample of n individuals the loglikelihood of the joint distribution of the items is given by

$$\log L = \sum_{h=1}^{n} \log f(\mathbf{v}_h)$$

EM algorithm

The EM algorithm presented here is given in Bartholomew (1987) for the logit model and for one latent variable. The method is easily extended to more than one latent variable but there are still technical problems to be solved especially when the number of latent variables exceeds two. Bartholomew described two different versions of the EM algorithm for fitting the model. The first version allow individuals to have any value of z and at the E-step of the algorithm we predict the value of z for each individual. In the second version the set of values of z is fixed and we have to predict how many individuals are located at each z. The second version is the one given in Bock and Aitkin (1981) and it can be applied to any type of response functions.

The second version of the algorithm, which is going to be used in this thesis, is now presented.

Suppose that z takes the values $z_1, z_2, \cdots, z_\nu$ with probabilities $h(z_1), h(z_2), \cdots, h(z_\nu)$. The marginal distribution is written:

$$f(\mathbf{v}_h) = \sum_{t=1}^{\nu} g(\mathbf{v}_h \mid z_t) h(z_t)$$

where

$$g(\mathbf{v}_h \mid z_t) = \prod_{i=1}^{s} \pi_i(z_t)^{v_{ih}} (1 - \pi_i(z_t))^{1 - v_{ih}}$$

We then have to maximize:

$$L = \sum_{h=1}^{n} \log f(\mathbf{v}_h)$$

By differentiating the log-likelihood respect to unknown parameters we get:

$$\frac{\partial L}{\partial \alpha_{il}} = \sum_{t=1}^{\nu} \frac{\partial \pi_i(z_t)}{\partial \alpha_{il}} \frac{\{r_{it} - N_t \pi_i(z_t)\}}{\pi_i(z_t)\{1 - \pi_i(z_t)\}} \qquad l = 0, 1. \tag{1.7}$$

Where,

$$\begin{aligned} r_{it} &= h(z_t) \sum_{h=1}^{n} v_{ih} g(\mathbf{v}_h \mid z_t) / f(\mathbf{v}_h) \\ &= \sum_{h=1}^{n} v_{ih} h(z_t \mid \mathbf{v}_h) \end{aligned} \tag{1.8}$$

and

$$\begin{aligned} N_t &= h(z_t)\sum_{h=1}^{n} g(\mathbf{v}_h \mid z_t)/f(\mathbf{v}_h) \\ &= \sum_{h=1}^{n} h(z_t \mid \mathbf{v}_h) \end{aligned} \tag{1.9}$$

The probability function $h(z_t \mid \mathbf{v}_h)$ is the probability that an individual h with response vector $\mathbf{v}_h$ is located at z_t.

The N_t could be interpreted as the expected number of individuals at z_t and r_{it} is the expected number of those predicted to be at z_t who will respond positively. The N_t and r_{it} are functions of the unknown parameters.

We define the steps of an EM algorithm as follows:

- **step1** Choose starting values for α_{i0} and α_{i1}

- **step2** Compute the values of r_{it} and N_t from (1.8) and (1.9)

- **step3** Obtain improved estimates of the parameters by solving (1.7)

- **step4** Return to step 2 and continue until convergence is attained.

Now if the response function is taken to be the logit equation (1.7) becomes:

$$\frac{\partial L}{\partial \alpha_{il}} = \sum_{t=1}^{\nu} z^l \{r_{it} - N_t \pi_i(z_t)\}, \qquad l = 0, 1. \tag{1.10}$$

There is a program called TWOMISS (Albanese and Knott 1992) which gives maximum likelihood estimates via this modified EM algorithm for the one and two latent trait model, using the logit model for the response function.

Bock and Aitkin (1981) proposed this method when z was in fact continuous. This is achieved by approximating the continuous z variables using Gauss-Hermite quadrature nodes available in FORTRAN libraries. They analyzed the data set Law School Aptitude Test which was also analyzed in Bock and Lieberman (1970).

The results they got from using 2 and 10 quadrature nodes are very close to the ones Bock and Lieberman (1970) obtained. They also fitted other types of prior distributions, a 10-point rectangular and a 10-point empirical prior distribution. The parameter estimates were close to the ones obtained with the normal prior distribution. They suggested that a number of quadrature nodes between 3 and 7 will be satisfactory for estimating a model with more than one latent variable. However, Shea (1984) show that many more nodes are needed in order to get a reasonable accuracy for the parameter estimates.

More than one latent variable

If there is more than one latent variable the above formulae require modification.

The response function takes the form

$$\text{logit}\pi_i(\mathbf{z}) = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j$$

The joint distribution of the manifest variables is given by

$$f(\mathbf{v}_h) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{v}_h \mid \mathbf{z}) h(\mathbf{z}) d\mathbf{z}$$

where the z's are assumed to be independent standard normal variables.

This probability can be approximated to any practical degree of accuracy by Gauss-Hermite quadrature,

$$f(\mathbf{v}_h) = \sum_{t_1=1}^{\nu_1} \cdots \sum_{t_q=1}^{\nu_q} g(\mathbf{v}_h \mid z_{1t_1}, \cdots z_{qt_q}) h(z_{1t_1}) \cdots h(z_{qt_q})$$

for $h = 1, \cdots, n$

where $z_{1t_1}, \cdots z_{qt_q}$ are tabled quadrature nodes and $h(z_{1t_1}) \cdots h(z_{qt_q})$ are the corresponding weights (Stroud and Secrest 1966).

The determination of the unknown parameters require the solution of $q + 1$

simultaneous non-linear equations for each item i.

The steps of the EM algorithm remain the same. The important thing that arises here when $q > 1$ is that there is no unique solution because of the fact that orthogonal transformations of the α_{ij}'s leave the value of the likelihood unchanged. The joint distribution of the manifest variables will remain unchanged after the transformation if the joint distribution of z^* (transformed) and z are the same. Constraints must be imposed on the parameters to give a unique solution.

Comments on the EM algorithm

As a criterion for the convergence of the EM algorithm we compare the relative change in the likelihood after each iteration with a very small number, (i.e. 0.00001).

The general theory of the EM algorithm, (Dempster et al. 1977), proves that each iteration of EM increases the likelihood and when the algorithm converges, it converges to a maximum of the likelihood.

The EM algorithm is simple to program and computationally efficient.

Problems that arise with the use of the EM algorithm in factor analysis are reported in Bentler and Tanaka (1983).

The drawbacks of the algorithm are 1) it does not check the second-order sufficiency conditions for a maximum, 2) it does not yield standard errors for the estimated parameters and 3) the convergence is slow when it reaches the maximum solution.

From practical experience it has been noticed that the EM may be sensitive to the starting values. So different starting values should be tried before reporting the ML solution.

The convergence properties of the EM algorithm have been studied by Wu (1983). He investigated conditions for which the EM algorithm converges to stationary points and to a unique maximum solution. He reported that if the likelihood function is unimodal and a certain differentiability condition is satisfied, then the EM algorithm converges to the unique maximum likelihood estimate.

Prior distribution

The distribution of the latent variable is arbitrary. What we mean by that is that we can transform the latent variable into another variable without actually changing the marginal distribution of the manifest variables. Usually the latent variable is assumed to follow a standard normal distribution but this is only for convenience, since any other distribution such as the uniform can be used, if the response function is modified suitably.

Bartholomew (1988), investigated empirically the effect of the change of the prior distribution, when a fixed response function is fitted, on the expected one- and two-way margins for models with one and more than one latent variables. The form of the prior distribution investigated is a symmetrical distribution with mean zero and variance one. He found that the expected one and two-way margins will not change much if the prior distribution is symmetrical and so the choice of the prior is a matter of convenience.

Other types of prior distributions have been looked at, such as the logistic, normal and rectangular. Similar results have been found.

As Bartholomew (1993) noted, in case where the latent variable is taken to have a normal distribution with mean μ and variance σ^2 there is an obvious location-scale transformation in the difficulty parameters. More complex differences in the distribution of the latent variables may be partly allowed for changing the parameters of the model. Tzamourani and Knott (1995) investigated this area using methods from robustness theory.

Sampling properties of the maximum likelihood estimates

From the first order asymptotic theory for maximum likelihood estimates we know that the maximum likelihood estimates have a sampling distribution which is asymptotically normal. Asymptotically, the sampling variances and covariances of the maximum likelihood estimates of the parameters (α_{ij}) are given by the elements of the inverse of the information matrix at the maximum likelihood solution. The inverse of the information matrix for a vector of parameters $\underline{\beta}$ is

$$[I(\hat{\beta})]^{-1} = -E\{\frac{\partial^2 \log L}{\partial\beta_j\partial\beta_k}\}_{\hat{\beta}=\beta}$$

Where,

$$\frac{\partial \log L}{\partial \beta_j} = \sum_{h=1}^{n} \frac{1}{f(\mathrm{v}_h)} \frac{\partial f(\mathrm{v}_h)}{\partial \beta_j}$$

$$\frac{\partial^2 \log L}{\partial\beta_j\partial\beta_k} = \sum_{h=1}^{n}\{\frac{1}{f(\mathrm{v}_h)}\frac{\partial^2 f(\mathrm{v}_h)}{\partial\beta_j\partial\beta_k} - \frac{1}{f(\mathrm{v}_h)^2}\frac{\partial f(\mathrm{v}_h)}{\partial\beta_j}\frac{\partial f(\mathrm{v}_h)}{\partial\beta_k}\}$$

By taking the expectation the first term vanishes

$$[I(\hat{\beta})]^{-1} = nE\{\frac{1}{f(\mathrm{v})^2}\frac{\partial f(\mathrm{v})}{\partial\beta_j}\frac{\partial f(\mathrm{v})}{\partial\beta_k}\}_{\hat{\beta}=\beta} \tag{1.11}$$

The expectation in equation (1.11) becomes

$$\sum_{\text{all v}}^{2^s} \frac{1}{f(\mathrm{v})}\frac{\partial f(\mathrm{v})}{\partial\beta_j}\frac{\partial f(\mathrm{v})}{\partial\beta_k}$$

In the program TWOMISS the standard errors of the maximum likelihood estimates are based on a approximation of equation (1.11) which is given by

$$I(\hat{\beta}) = \{\sum_{h=1}^{n}\frac{1}{f(\mathrm{v}_h)^2}\frac{\partial f(\mathrm{v}_h)}{\partial\beta_j}\frac{\partial f(\mathrm{v}_h)}{\partial\beta_k}\}^{-1}$$

In our analysis it is often the case that some of the estimated parameters take large values. In cases like these asymptotic standard errors are not trustworthy since the sampling distribution of these parameters can be skewed or a mixture of two distributions. Other techniques can be used for estimating the standard errors such as jackknife and bootstrapping. Albanese and Knott (1994) have investigated the behaviour of the standard errors of the MLE for the one factor logit/probit model by using bootstrapping methods. They reported that when the discrimination parameters are small the asymptotic theory works well, but when they get large it can be inadequate. It has also been observed that in the one- and two-factor logit/probit model the standard errors of large discriminating parameters are large as well.

Goodness of fit

A test of the latent trait model may be based on either the Pearson goodness of fit (X^2), or the likelihood ratio statistic (G^2). If we denote by s the number of manifest variables and by T the number of all the possible response patterns (2^s)

$$X^2 = \sum_{i=1}^{T} \frac{(O_i - E_i)^2}{E_i}$$

$$G^2 = -2 \sum_{i=1}^{T} O_i \log \frac{E_i}{O_i}$$

where O_i is the observed frequency of the response pattern i and E_i is the expected frequency of the response pattern i (estimating from the model).

The degrees of freedom for each statistic are $2^s - 2s(q+1) - 1$, where q is the number of latent variables in the model.

As the number of the manifest variables increases the number of response patterns increases as well. If the sample size n is kept constant or if it is small compared to 2^s, the frequencies of the response patterns will become small. It is well known that when a large number of expected frequencies are too low, the χ^2 approximation for the distribution of X^2 and G^2 may not be valid. Informal techniques that are sometimes used to overcome the problem of cells with low or zero frequencies include combining cells with low expected frequencies and subtracting one degree of freedom each time a grouping is made or adding a small constant to each cell. The grouping technique is not successful in cases when the sparseness is very severe and there are times when there are no degrees of freedom left to carry on the test. The second technique has the disadvantage that it increases the sample size when again the sparseness is severe.

An additional check of the model can be made by comparing the observed and expected frequencies of the one- two- and three- way marginal frequencies. From these tables we can detect cases where the model does not fit well. This way for checking the fit of the model has been used in de Menezes and Bartholomew (1996).

Reiser and VandenBerg (1994) used some simulation results and suggested that when the number of manifest variables is greater than 8 the goodness-of-fit must be checked from the one- and two- way margins which are obtained in their paper by following the limited-information method for factor analysis of dichotomous variables where the response information is used only from the first- and second-order marginal distributions, (Christofferson 1975; Muthén 1978; Muthén 1984). The power of that test is not affected by sparseness but at the same time Type I error becomes quite high for nine or ten variables and so more investigation is needed.

In the same study it was reported that when we use the full information method the Pearson statistic is more resistant than the likelihood ratio to the effect of sparseness for up to 7 manifest variables.

1.4 Factor analysis for mixed manifest variables

As already mentioned in the introduction of this chapter, this thesis will deal with the development of a latent variable model for binary and continuous manifest variables. The approach which will be used in this thesis is an extension of Bartholomew's work for fitting latent variable models. The new developments will be discussed in the next chapter.

In the literature the underlying variable approach has been used for the mixed case. Contributions have been made by Muthén, Jöreskog and Sörbom, and Arminger and Küsters. Their work cover a wide range of models which also allow relationships among the latent variables and inclusion of exogenous (explanatory) variables. Their approaches will be discussed here.

Muthén (1984) proposed a three stage estimation method which is actually an extension of the theory developed by Muthén and Christoffersson (1981) and Muthén (1978) for categorical manifest variables.

At the first stage first order statistics such as thresholds, means and variances are estimated by ML, in the second stage, second order statistics such as tetrachoric, polychoric (Olsson 1979) and polyserial (Olsson, Drasgow, and Dorans 1982) correlations are estimated by conditional ML for given first stage estimates.

Olsson (1979) pointed out that bad estimates of the tetrachoric and polychoric correlations can be obtained when some expected cell frequencies are low and also that the standard errors of the estimates are not reasonable when the expected cell frequencies are less than 10.

At the third stage the parameters of the structural part of the model are estimated using a limited-information generalized least squares method. The estimation at the first two stages is based on maximizing the univariate and bivariate log likelihood function for the latent response variables (underlying variables). That is the reason why the ML estimated parameters are called limited information ML estimates. With this method they obtain a consistent estimator of the asymptotic covariance matrix of the estimates of the first two stages.

The partition of the ML estimation might have an effect at the goodness of fit and the statistical properties of the estimates derived at the third stage of the estimation procedure. However Muthén (1984) claims that his estimates will be always asymptotically normally distributed and efficient.

Although Muthén (1984) promises that his estimation procedure provides large sample chi-square tests of fit and standard errors of the parameter estimates for the mixed case as well, his method can only give covariances among polychoric correlations as an extension of Olsson (1979) work which derives only variances but there is nothing mentioned for covariances for polyserial correlations. This cannot be clarified with his examples because they both refer to ordinal and dichotomous scale variables.

However his method covers a wide range of structural models (with exogenous variables as well) for the case of metric manifest variables. These models can be estimated using a program LISCOMP, (Muthén 1987). In the LISCOMP manual he reports that a simplified weight matrix is available which seems to work well and demands less computing time and memory. He also reports that the GLS estimators that use the full weight matrix require a lot of computing time because the weight matrix grows very rapidly with the increase of the manifest variables.

There is a series of papers (Browne 1974,1982, 1984) describing the GLS estimators in the analysis of covariance structures.

Lee, Poon, and Bentler (1992) describe a two-stage procedure for analysing structural equation models with continuous and polytomous items. At the first stage they estimate thresholds, polychoric, polyserial covariances using a full maximum likelihood estimation and at the second stage the parameters of the structural part of the model are estimated using generalized least squares method.

By estimating all the parameters simultaneously using ML approach (see Poon and Lee 1987) the final estimates have ML properties such as consistency, asymptotic efficiency and normality. As a consequence of the ML estimates at the first stage is that the weight matrix which is involved at the second stage will be more accurate. The estimates follow an asymptotic normal distribution and so their covariance matrix is obtained from the inverse of the information matrix.

Their method is similar to Muthén (1984) method. The difference is actually at the estimation of the parameters at the first stage. They use a full ML estimation and Muthén uses a limited information ML estimation. Muthén's method looks less computationally heavy although it does not guarantee that the joint distribution of the parameters is asymptotically multivariate normal. The first stage estimates are obtained by the iterative Fletcher-Powell algorithm and the second stage estimates by the iterative Gauss-Newton algorithm.

They did a simulation study for comparing their estimation procedure with LISCOMP approach. In brief, they commented that at sample sizes 100 and 200, LISCOMP gave better parameter estimates than Lee et al. 1992. However LISCOMP gave goodness-of-fit statistics that were not chi-squared distributed. Lee et al. (1992) claim that the fact that Muthén is not using a full ML estimation at the first stage will have a bad effect on the inference of the model at the second stage of GLS estimation. Their program is not available and so comparisons with our approach is not possible. Because of the full ML estimation at the first stage their method is limited to a small number of categorical variables which in the paper is mentioned to be less than four.

All these methods require the estimation of a weight matrix. This matrix grows very much with increase of the number of manifest variables and even more when we analyze mixed items. In practice that limits the number of variables for analysis to 15-20.

The LISREL (Jöreskog and Sörbom 1993b) approach is also based on the analysis of polychoric and polyserial correlations (estimated using PRELIS) and a weighted least squares method for estimating the structural parameters. Jöreskog (1990), among other new features of LISREL he mentions that polyserial correlation coefficients can be obtained from bivariate summary statistics consisting of the frequency in each cell, the mean and the variance of the continuous variables in each category of the categorical items.

For the case where exogenous variables are included in the model the correlation coefficients estimated by PRELIS are unconditional that means that assumptions for the normality of the underlying and the exogenous variables are required. On the contrary LISCOMP estimates correlation coefficients of the underlying variables conditional on the exogenous variables that means that only the normality assumption of the underlying variable given the exogenous variables is required and this is probably preferred.

One of the estimation procedures for the structural parameters in LISREL is called weighted least squares (WLS), (see Jöreskog and Sörbom 1988). The difference between GLS and WLS is that the first method requires normality of the response variables and the latter is asymptotically distribution free. The difference is in the weight matrix used. The weight matrix in WLS requires the computation of fourth-order central moments, (that requires large sample sizes), but it gives correct asymptotic chi-squares and standard errors. It looks as if WLS estimation method is more appropriate for the mixed items case, although it appears to be computational heavy as the number of variables increases.

So far the methods which are presented here can fit a latent variable model on mixed items by treating all items as metric and by using GLS or WLS as estimation method.

Arminger and Küsters (1988) have also adopted an underlying variable approach in which all the observed variables are treated as metric variables but in which the estimation method is maximum likelihood.

They give a very general framework for estimating simultaneous equation models, (endogenous observed variables connected to latent endogenous variables), with observed variables of level of measurement of any type and metric latent variables.

Their formulation allows metric and dummy (0/1) exogenous variables to be included in the model.

In their formulation there are four different type of variables namely the endogenous observed variables (x), the endogenous latent variable (z), the underlying response variable (x^*), and the exogenous variables (ξ).

They distinguish between a single indicator case in which the latent variable is equal to an endogenous variable and the multiple indicator case in which several endogenous observed variables are connected to only one latent variable. The vector of $\mathbf{x}$ can be partitioned into subsets, each of which depends only on one latent variable. These sets of variables seem to be treated separately in the estimation because the loglikelihood function is maximized for each of this sets. The set up of the model is the same as LISREL model, (confirmatory factor analysis).

Only the multiple indicator case will be reviewed here, since we cannot see the use of the single indicator case. Also because we are not interested in the structural part of the model, relations between latent variables and between latent variables and exogenous variables, more emphasis in the presentation of their method will be given to the part related to our work that is the measurement relations rather than the structural part of the model.

There are three different type of relationships to be defined in the analysis. First, the metric endogenous latent variable z modelled as in an ordinary structural equation model, here also depends on exogenous variables ξ. Second, each observed variable (x) is related to an underlying variable (x^*) via threshold models, (measurement relations). Third, a set of underlying variables $\mathbf{x}^*$ is related to one and only one latent variable (z) via a linear factor model.

Arminger and Küsters (1989) start their analysis by defining the marginal distribution for each observed item, see also Bartholomew (1987). Because of the existence of exogenous variables the marginal distribution is conditional on these exogenous variables and for one latent variable is written:

$$f(\mathbf{x} \mid \xi) = \int_{R(z)} g(\mathbf{x} \mid z, \xi) h(z \mid \xi) dz \qquad (1.12)$$

where $R(z)$ denotes the domain of z.

We need to define the conditional distribution $g(\mathbf{x} \mid z, \xi)$ for each observed variable depending on its level of measurement. The latent variable z is taken continuous and so only latent trait models are included.

For a random sample of n individuals the loglikelihood function to be maximized is the sum with respect to all individuals of the logarithm of the marginal density given in equation (1.12).

Three assumptions/simplifications are required to be made for the estimation of the model:

1. Conditional independence is assumed within the elements of each set and between sets. So if there are l sets of I_l elements in each sets:

$$g(\mathbf{x} \mid \mathbf{z}) = \prod_{j=1}^{l} \prod_{i=1}^{I_l} g(x_{j,i} \mid z_j)$$

2. The endogenous observed variables $\mathbf{x}$ depends on the exogenous variables ξ only through the latent variables $\mathbf{z}$.

3. They assume that the vector $\mathbf{x}$ can be partitioned into I subsets, each depending on only one latent variable z. In other words this is the simple structure principle which assumes that each observed variable x is connected only to one latent variable.

As already mentioned above the underlying response variables are connected with the latent variables through a linear factor model:

$$x^*_{j,i} = \gamma_{j,i} + \lambda_{j,i} z_j + \epsilon_{j,i} \tag{1.13}$$

where j denotes the subset number and i is the index denoting the different endogenous observed variables in each subset, and $\epsilon_{j,i}$ is the error term assumed to be independent of z_j.

All the endogenous observed variables (metric, categorical) are modelled using equation (1.13).

The specification of the measurement relations between $x_{j,i}$ and $x^*_{j,i}$ and the error term $\epsilon_{j,i}$ define the form of the conditional distribution $g(\mathbf{x}_{j,i} \mid z_j)$. The measurement relations are given below.

For the case where the x variable is metric the measurement model to be used is:

$$x_i = x_i^*$$

where $\epsilon_{j,i} \sim N(0, \Psi_{j,i})$ and the conditional density is:

$$x_{j,i} \mid z_j \sim N(\gamma_{j,i} + \lambda_{j,i} z_j, \Psi_{j,i})$$

For the case where the x variable is ordinal with $c_{j,i}$ categories the threshold model to be used is:

$$x_{j,i} = k \quad iff \quad \tau_{j,i,k-1} < x^*_{j,i} \leq \tau_{j,i,k}, \quad k = 1, \cdots, c_{j,i}$$

with

$$-\infty = \tau_{j,i,0} < \tau_{j,i,1} \cdots < \tau_{j,i,c_{j,i}} = +\infty$$

and $\epsilon \sim N(0,1)$, the conditional density function is:

$$g(x_{j,i} = k \mid z_j) = \int_{\tau_{j,i,k-1}}^{\tau_{j,i,k}} \phi(x^*) dx^*$$

where $\phi(x^*)$ is a function of x^* following $N(\gamma_{j,i} + \lambda_{j,i} z_j, 1)$.

For the case where the x variable is unordered categorical with $c_{j,i}$ categories the threshold model to be used:

$$x_{j,i} = k \quad iff \quad x^*_{j,i,k} > x^*_{j,i,l} \quad \text{for} \quad l = 1, \cdots, c_{j,i}, \quad l \neq k$$

the error terms $\epsilon_{j,i,k}$ are independent identically distributed with the extreme value distribution with density function:

$$F(\epsilon_{j,i,k}) = \exp[-\exp(-\epsilon_{j,i,k})]$$

giving the multinomial logit.

The conditional density function is:

$$g(x_{j,i} = k \mid z_j) = \frac{\exp(\gamma_{j,i,k} + \lambda_{j,i,k} z_j)}{\sum_{l=1}^{c_{j,i}} 1 + \exp(\gamma_{j,i,l} + \lambda_{j,i,l} z_j)} \tag{1.14}$$

Measurement relations are also defined for censored metric variables.

The model parameters contained in the conditional distribution $x \mid z$ together with the asymptotic covariance matrix are estimated by a limited marginal likelihood approach. The structural parameters which connect the latent variables with the exogenous variables together with their asymptotic covariance matrix are estimated using the weighted or unweighted version of Amemiya's principle. More can be found in Arminger and Küsters (1988).

The limited marginal likelihood approach refers to the maximization of the marginal distribution given in equation (1.12) for each multiple indicator set. That means that a one factor model is fitted on a subset of the observed endogenous variables. The first derivatives of the loglikelihood with respect to the unknown parameters are given in Arminger and Küsters (1988) for each type of observed variable. A consistent estimate of the asymptotic covariance matrix of the parameter estimates is computed using the inverse of an approximation of the information matrix.

An EM algorithm is suggested for the maximization of the loglikelihood function which requires the maximization of the expected value of the logarithm of the complete data likelihood given the observed data and the parameters estimated from the previous iteration:

$$\sum_{h=1}^{n} E\{\ln g(\mathrm{x}, z \mid \xi, \theta^{d+1}) \mid \mathrm{x}, \theta^{d}\}$$

where d denotes the dth iteration and θ is a vector with the unknown parameters. This maximization is achieved sequentially for each observed variable x_i , $i = 1, \cdots, I$

Results from this method have been presented in their paper Arminger and Küsters (1989) for the case of three endogenous observed variables, one metric exogenous and one latent variable using GAUSS routines. Arminger and Küsters (1988) theory has not been implemented in any software such as MECOSA, (Schepers and Arminger 1992).

The differences and similarities with our approach will be discussed in detail in Chapter 2 where our approach will be presented.

We shall compare our approach for handling mixed items with the statistical software LISCOMP in the chapter with the applications (Chapter 4).

1.5 Scaling methods

Social scientists are particularly interested in locating individuals on the dimensions of the latent factor space according to their response patterns. The latent scores can be substituted for the manifest variables in analysis with other independent variables of interest.

Scoring methods have been proposed in the literature for the known latent variable models.

For binary responses the total score of each individual which is obtained by adding the answers of all s items provides a simple scoring method which gives the same weight to all the items.

Bartholomew (1980) proposed a method for scaling a set of binary responses using the logit factor model and in Bartholomew (1981) that method was extended to the factor model with continuous responses. He argues that as latent variables in the model are random, Bayes' theorem provides the logical link between the data and the latent variables. Hence, the mean of the posterior distribution of z given $\mathbf{v}$, $(E(z \mid \mathbf{v}))$ can be used to score $\mathbf{v}$. The advantage of using the posterior mean as a scaling method is that it is approximately a linear function of the components $V = \sum_{i=1}^{s} \alpha_{i1} v_i$ if the α_{i1} coefficients are small and for any prior distribution.

Knott and Albanese (1993) investigated $h(z \mid \mathbf{v})$ for the logistic latent trait model for binary responses. They proved that if the conditional distribution of z when all responses are zero is normal, then the conditional distribution of z for any set of responses is normal. They also comment that this result is not altered if some of the alpha coefficients are large and the number of items increases.

An alternative method provides component scores proposed by Bartholomew (1984b) that method avoids the calculation of the posterior mean and the numerical integrations involved. In that paper he investigated the logistic latent model for binary responses where the latent variable z follows a uniform distribution on (0,1). From the posterior distribution of the latent variable given the observed response pattern it is clear that the posterior distribution depends on v only through V; V is thus a Bayesian sufficient statistic for z. The sufficiency of V was noted by Birnbaum in Lord and Novick (1968), (ch 18), for a fixed effects version of the model. The sufficiency depends on the choice of the response function, it holds for the logit but not for the probit.

The component score has an obvious intuitive appeal because of its linearity and the fact that it weights the manifest variables in proportion to their contribution to the common factor.

Bartholomew (1984b) shows than an approximation can be obtained:

$$E(Y \mid \mathrm{v}) \approx (1+V)/(2+A)$$

where

$V = \sum_{i=1}^{s} \alpha_{i1} v_i$ and $A = \sum_{i=1}^{s} \alpha_{i1}$

This result is exact if $\pi_i = 1/2$ and $\alpha_{i1} = 1$ for all i.

The calculations suggest that $E(y \mid \mathrm{v})$ and V are almost equivalent for scaling purposes. This result depends on the choice of uniform prior distribution for y.

Bartholomew (1984b) and Knott and Albanese (1993) have shown that for the one logit/logit model and the one logit/probit model for binary responses both scaling methods give the same ranking to response patterns/individuals.

Analogous results have been derived for the linear factor model. Bartholomew (1984a) shows that the component scores are sufficient statistics given conditional independence for the items and that the posterior distribution of $w_i \mid z$ is of exponential type. By applying his results to a special case which is the linear factor model for which we assume that $g(w_i \mid z) \sim N(\mu, \sigma^2)$, we obtain the component score for the linear factor model to be $\sum_{i=1}^{r} \frac{\lambda_{ij}}{\Psi_{ii}} w_i$.

1.6 Outline of the thesis

This chapter is an overview of the existing approaches in the literature of latent variable models for binary, metric and mixed manifest variables. The remainder of the thesis is organized as follows.

In Chapter 2 a latent trait model (continuous factor space) is developed for fitting mixed, (binary and metric), manifest variables. We discuss the estimation method of the model parameters, standard errors, goodness-of-fit and scoring methods for the individuals on the latent factor space.

In Chapter 3 a latent class model (discrete factor space) is developed for fitting mixed, (binary and metric), manifest variables. We discuss the estimation method of the model parameters, standard errors and the allocation of individuals in the latent classes fitted in the model.

Two pieces of software have been developed for fitting the latent trait and the latent class model to mixed manifest variables. The models developed are fitted in four data sets vary in number of cases and number of manifest variables. The results of the analysis are given in Chapter 4.

In Chapter 5 the latent trait model developed in Chapter 2 is extended to handle incomplete data. We discuss the set up of the model and the estimation method. A number of applications are presented to illustrate the use of the model and the information that can be obtained about attitude from non-response.

In Chapter 6 the results presented in Chapter 2 are put in a general framework that can handle manifest variables with conditional distributions in the exponential family. That general framework allows for a common estimation method and a generalization of the results derived in Chapter 2.

Finally, Chapter 7 concludes with an overview of the contribution of the current research and proposals for future research.

Chapter 2

Latent trait model

2.1 Introduction

Using the existing theory of latent variable models in the form adopted by Bartholomew (1987), an analogous technique is presented here for cases where the manifest variables are of mixed type. Some of the manifest variables are binary and some are continuous. For the continuous part the linear factor model is used and for the binary part the response function approach is followed. Both these models are described in Chapter 1, sections 1.2 and 1.3.2 respectively.

The mixed model allows a single analysis of the binary and the continuous part. The latent variables are assumed to have continuous independent variables with standard normal distributions.

2.2 Latent trait model with mixed manifest variables

Suppose there are p manifest variables where r are continuous and s are discrete, $(r + s = p)$. The continuous manifest variables are denoted by $\mathbf{w}$ and the binary variables are denoted by $\mathbf{v}$. Let us suppose that their relationship is accounted for by a number of q continuous variables $\mathbf{y}$.

As already mentioned in Chapter 1 (section 1.3.2), the formulation of the model starts with the joint distribution of the manifest variables because this is the one we observe.

$$f(\mathbf{x}) = \int_{R_y} g(\mathbf{x} \mid \mathbf{y}) h(\mathbf{y}) d\mathbf{y}$$

where R_y is the range space of $\mathbf{y}$.

Under the assumption of conditional independence:

$$g(\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{p} g(x_i \mid \mathbf{y})$$

We have to decide about the form of the conditional distributions $g(x_i \mid \mathbf{y})$ and the form of the prior distribution of the latent variables, $h(\mathbf{y})$. We have already seen in Chapter 1 (section 1.3.2) that the form of the prior distribution is quite arbitrary. The same could be assumed for the conditional distributions, $g(x_i \mid \mathbf{y})$, since the vector $\mathbf{y}$ is not observed and so it cannot be held fixed. However, Bartholomew's approach sets some restrictions on the choice of these conditional distributions as it will be shown now.

Our first interest is to pass from the $p = (r+s)$ manifest variables to q unobserved variables where q is much less than p. The $\mathbf{x}$'s contain all the information about $\mathbf{y}$. Bartholomew took that one step further by saying that it will be desirable if summary statistics can be found to contain the information about $\mathbf{y}$, which these observable summary statistics will be of q-dimension rather than p.

Thus the problem becomes what form the conditional distributions, $g(x_i \mid \mathbf{y})$, must have in order the $h(\mathbf{y} \mid \mathbf{x})$ to depend on $\mathbf{x}$ through a q function of $\mathbf{x}$. That is what called by Bartholomew (1987), the *sufficiency principle.* This principle is fundamental in the approach we use in this thesis.

Barankin and Maitra (1963) have given the necessary and sufficient conditions in order the sufficiency principle to be satisfied which says that at least $p - q$ of the conditional distributions $g(x_i \mid \mathbf{y})$ must be of exponential type defined as:

$$g(x_i \mid \mathbf{y}) = F_i(x_i) G_i(\mathbf{y}) \exp \sum_{j=1}^{q} u_{ij}(x_i) \phi_j(\mathbf{y})$$

The posterior distribution of $h(\mathbf{y} \mid \mathbf{x})$ is:

$$h(\mathbf{y} \mid \mathbf{x}) = \frac{h(\mathbf{y}) \prod_{i=1}^{p} F_i(x_i) G_i(\mathbf{y}) \exp \sum_{j=1}^{q} X_j \phi_j(\mathbf{y})}{\int h(\mathbf{y}) \prod_{i=1}^{p} F_i(x_i) G_i(\mathbf{y}) \exp \sum_{j=1}^{q} X_j \phi_j(\mathbf{y})} \tag{2.1}$$

where $X_j = \sum_{i=1}^{p} u_{ij}(x_i)$ and $\phi_j(\mathbf{y})$ is a function of $\mathbf{y}$.

From equation (2.1) we see that the product $\prod_{i=1}^{p} F_i(x_i)$ cancels out and we left with the posterior distribution $h(\mathbf{y} \mid \mathbf{x})$ to depend on $\mathbf{x}$ only through the q components X_j. This X_j is sufficient for $\mathbf{y}$.

For the binary case the sufficient statistic X_j is proven to be, (see Bartholomew 1984b):

$$X_j = \sum_{i=1}^{s} \alpha_{ij} v_i$$

and for the continuous case, (see Bartholomew 1984a), where the linear factor model is used it is:

$$X_j = \sum_{i=1}^{r} \frac{\lambda_{ij}}{\Psi_{ii}} w_i$$

Now for the case of binary and continuous items because both the conditional distributions for the binary and the continuous items belong to the exponential family, the exponents in the two parts are added up and so the component is:

$$X_j = \sum_{i=1}^{s} \alpha_{ij} v_i + \sum_{i=1}^{r} \frac{\lambda_{ij}}{\Psi_{ii}} w_i$$

We will come back to these results when we discuss scoring methods.

The choice of the distribution of the latent variable y is arbitrary. In this thesis is taken to be the standard normal because it leads to linear models, so:

$$z_j = \phi_j(y), \qquad j = 1, \cdots, q$$

The first use of the sufficiency principle is that all the information we need to know for z is contained in the sufficient statistic X which for the case of q latent variables is a q function of x.

The second one is that the sufficient statistic X for the binary and the continuous case is linear in the x's and that can be used to construct measures for scaling individuals on the latent factor space by also allowing for a different weight to be given to each item.

Now all the x's are not constrained to have the same type of conditional distribution. So for the case of mixed type of manifest variables the binary items will have the Bernoulli distribution and the metric variables will have the normal.

2.2.1 One factor latent trait model with mixed manifest variables

The model will be presented here for one latent variable and it will be extended later to the case of more than one latent variable. We denote the conditional distribution of the manifest variables by $g(w_i \mid z)$ and $g(v_i \mid z)$ for continuous and binary variables respectively.

Under the assumption of conditional independence,

$$g(\mathbf{w} \mid z) = \prod_{i=1}^{r} g(w_i \mid z)$$

$$g(\mathbf{v} \mid z) = \prod_{i=1}^{s} g(v_i \mid z)$$

$$g(\mathbf{x} \mid z) = \prod_{i=1}^{r} g(w_i \mid z) \prod_{i=1}^{s} g(v_i \mid z)$$

The joint distribution of the manifest variables is given by:

$$f(\mathbf{x}_h) = \int_{-\infty}^{\infty} g(\mathbf{w}_h \mid z) g(\mathbf{v}_h \mid z) h(z) dz \qquad (2.2)$$

where $\mathbf{x}_h$ represents the responses to the p manifest variables of the h^{th} individual and $h(z)$ is the prior distribution of the latent variable, assumed to be standard normal.

We want to examine if $f(\mathbf{x}_h)$ is an adequate representation of the data for a single latent variable z.

Using the sufficiency principle described above the form of the conditional distributions for the continuous and the binary items are taken from the exponential family.

The continuous manifest variables are fitted using the linear factor model described in Chapter 1 (section 1.2). So the conditional distribution of $w_i \mid z$ is given by:

$$g(w_i \mid z) = (2\pi)^{-1/2}\Psi_{ii}^{-1/2}\exp(-\frac{1}{2\Psi_{ii}}(w_i - \mu_i - \lambda_{i1}z)^2) \tag{2.3}$$

The parameter λ_{i1} for the one factor model is abbreviated with λ_i.

The binary manifest items are fitted using the response function method, also described in Chapter 1 (section 1.3.2). The conditional distribution of $v_i \mid z$ is given by:

$$g(v_i \mid z) = \pi_i(z)^{v_i}(1 - \pi_i(z))^{1-v_i} \tag{2.4}$$

Where the response function takes the form:

$$\text{logit}\pi_i(z) = \alpha_{i0} + \alpha_{i1}z \tag{2.5}$$

The estimation of the parameters, (α_{i0}, α_{i1}, μ_i, λ_i and Ψ_{ii}), is based on the marginal distribution of the manifest variables given by equation (2.2).

This probability can be approximated to any practical degree of accuracy by Gauss-Hermite quadrature, i.e.,

$$f(\mathbf{x}_h) = \sum_{t=1}^{\nu} g(\mathbf{w}_h \mid z_t)g(\mathbf{v}_h \mid z_t)h(z_t) \quad h = 1, \cdots, n \tag{2.6}$$

where z_t is a tabled quadrature node and $h(z_t)$ is the corresponding weight (Stroud and Secrest 1966).

This method involves choosing the number and location of the nodes $z_1, z_2, \cdots, z_\nu$, so that the various sums over t approximate the corresponding integrals that arise in the continuous time treatment. In order to estimate the unknown parameters for the discrete and the continuous part we will maximize the log-likelihood function. The maximization of the likelihood will be achieved using optimization routines more specifically an E-M algorithm.

The log-likelihood for a random sample of size n is

$$L = \sum_{h=1}^{n} \log f(\mathbf{x}_h)$$

or

$$L = \sum_{h=1}^{n} \log f(\mathbf{x}_h) = \sum_{h=1}^{n} \log \int_{-\infty}^{\infty} g(\mathbf{w}_h \mid z) g(\mathbf{v}_h \mid z) h(z) dz$$

Finding partial derivatives, we have

$$\begin{aligned} \frac{\partial L}{\partial \alpha_{il}} &= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \frac{\partial f(\mathbf{x}_h)}{\partial \alpha_{il}} \\ &= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \sum_{t=1}^{\nu} h(z_t) g(\mathbf{w}_h \mid z_t) \frac{\partial g(\mathbf{v}_h \mid z_t)}{\partial \alpha_{il}} \end{aligned} \tag{2.7}$$

where,

$$\begin{aligned} \frac{\partial g(\mathbf{v}_h \mid z_t)}{\partial \alpha_{il}} &= \frac{\partial}{\partial \alpha_{il}} [\prod_{i=1}^{s} \{\pi_i(z_t)\}^{v_{ih}} \{1 - \pi_i(z_t)\}^{1-v_{ih}}] \\ &= g(\mathbf{v}_h \mid z_t) \{ \frac{v_{ih}}{\pi_i(z_t)} - \frac{(1 - v_{ih})}{(1 - \pi_i(z_t))} \} \frac{\partial \pi_i(z_t)}{\partial \alpha_{il}} \end{aligned} \tag{2.8}$$

$$i = 1, 2, ..., s; \quad l = 0, 1.$$

Hence, by substituting (2.8) in (2.7) and interchanging the summations, we find

$$\frac{\partial L}{\partial \alpha_{il}} = \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \sum_{t=1}^{\nu} h(z_t) g(\mathbf{w}_h \mid z_t) g(\mathbf{v}_h \mid z_t) \{ \frac{v_{ih}}{\pi_i(z_t)} - \frac{(1 - v_{ih})}{(1 - \pi_i(z_t))} \} \frac{\partial \pi_i(z_t)}{\partial \alpha_{il}}$$

$$
\begin{aligned}
&= \sum_{t=1}^{\nu} h(z_t)\frac{\partial \pi_i(z_t)}{\partial \alpha_{il}}\{\sum_{h=1}^{n} \frac{g(\mathbf{w}_h \mid z_t) g(\mathbf{v}_h \mid z_t)}{f(\mathbf{x}_h)}[\frac{v_{ih} - \pi_i(z_t)}{\pi_i(z_t)\{1-\pi_i(z_t)\}}]\} \\
&= \sum_{t=1}^{\nu} h(z_t)\frac{\partial \pi_i(z_t)}{\partial \alpha_{il}} \times \\
&\quad \frac{\sum_{h=1}^{n} g(\mathbf{w}_h \mid z_t) g(\mathbf{v}_h \mid z_t) v_{ih}/f(\mathbf{x}_h) - \sum_{h=1}^{n} g(\mathbf{w}_h \mid z_t) g(\mathbf{v}_h \mid z_t)\pi_i(z_t)/f(\mathbf{x}_h)}{\pi_i(z_t)\{1-\pi_i(z_t)\}}.
\end{aligned} \tag{2.9}
$$

Hence, equation (2.9) can be written:

$$
\frac{\partial L}{\partial \alpha_{il}} = \sum_{t=1}^{\nu} \frac{\partial \pi_i(z_t)}{\partial \alpha_{il}} \frac{\{r_{1it} - N_t \pi_i(z_t)\}}{\pi_i(z_t)\{1-\pi_i(z_t)\}} \tag{2.10}
$$

where, r_{1it} and N_t are defined in (2.12) and (2.13).

Finally, for the response function defined in equation (2.5) the first derivative of the loglikelihood respect to α_{il} parameters becomes:

$$
\frac{\partial L}{\partial \alpha_{il}} = \sum_{t=1}^{\nu} z^l \{r_{1it} - N_t \pi_i(z_t)\}, \qquad l = 0, 1 \tag{2.11}
$$

where,

$$
\begin{aligned}
r_{1it} &= h(z_t) \sum_{h=1}^{n} v_{ih} g(\mathbf{w}_h \mid z_t) g(\mathbf{v}_h \mid z_t)/f(\mathbf{x}_h) \\
&= \sum_{h=1}^{n} v_{ih} h(z_t \mid \mathbf{x}_h)
\end{aligned} \tag{2.12}
$$

and

$$
\begin{aligned}
N_t &= h(z_t) \sum_{h=1}^{n} g(\mathbf{w}_h \mid z_t) g(\mathbf{v}_h \mid z_t)/f(\mathbf{x}_h) \\
&= \sum_{h=1}^{n} h(z_t \mid \mathbf{x}_h)
\end{aligned} \tag{2.13}
$$

The probability function $h(z_t \mid \mathbf{x}_h)$ is the probability that an individual h with response vector $\mathbf{x}_h$ is located at z_t.

The N_t could be interpreted as the expected number of individuals at z_t and the r_{1it} could be interpreted as the expected number of individuals at z_t that have responded positively to binary item v_i.

We carry on by computing the partial derivatives for the parameters of the continuous part.

$$\begin{aligned}
\frac{\partial L}{\partial \lambda_i} &= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \frac{\partial f(\mathbf{x}_h)}{\partial \lambda_i} \\
&= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \sum_{t=1}^{\nu} h(z_t) g(\mathbf{v}_h \mid z_t) \frac{\partial g(\mathbf{w}_h \mid z_t)}{\partial \lambda_i} \qquad (2.14)
\end{aligned}$$

Where,

$$\begin{aligned}
\frac{\partial g(\mathbf{w}_h \mid z_t)}{\partial \lambda_i} &= \frac{\partial}{\partial \lambda_i}[\prod_{i=1}^{r}(2\pi)^{-1/2}\Psi_{ii}^{-1/2}\exp(-\frac{1}{2\Psi_{ii}}(w_{ih}-\mu_i-\lambda_i z_t)^2)] \\
&= \frac{\partial}{\partial \lambda_i}[(2\pi)^{-1/2}\Psi_{11}^{-1/2}\exp(-\frac{1}{2\Psi_{11}}(w_{1h}-\mu_1-\lambda_1 z_t)^2) \times \cdots \\
&\quad \times(2\pi)^{-1/2}\Psi_{rr}^{-1/2}\exp(-\frac{1}{2\Psi_{rr}}(w_{rh}-\mu_r-\lambda_r z_t)^2)] \\
&= (2\pi)^{-1/2}\Psi_{11}^{-1/2}\exp(-\frac{1}{2\Psi_{11}}(w_{1h}-\mu_1-\lambda_1 z_t)^2) \times \cdots \\
&\quad \times(2\pi)^{-1/2}\Psi_{rr}^{-1/2}\exp(-\frac{1}{2\Psi_{rr}}(w_{rh}-\mu_r-\lambda_r z_t)^2)(2\pi)^{-1/2}\Psi_{ii}^{-1/2} \\
&\quad \frac{\partial}{\partial \lambda_i}\exp(-\frac{1}{2\Psi_{ii}}(w_{ih}-\mu_i-\lambda_i z_t)^2) \qquad (2.15)
\end{aligned}$$

where,

$$\begin{aligned}
\frac{\partial}{\partial \lambda_i}\exp(-\frac{1}{2\Psi_{ii}}(w_{ih}-\mu_i-\lambda_i z_t)^2) &= \exp(-\frac{1}{2\Psi_{ii}}(w_{ih}-\mu_i-\lambda_i z_t)^2) \times \\
&\quad \frac{z_t}{\Psi_{ii}}(w_{ih}-\mu_i-\lambda_i z_t) \qquad (2.16)
\end{aligned}$$

by substituting (2.16) to (2.15) we get:

$$\frac{\partial g(\mathbf{w}_h \mid z_t)}{\partial \lambda_i} = g(\mathbf{w}_h \mid z_t)[\frac{z_t}{\Psi_{ii}}(w_{ih}-\mu_i-\lambda_i z_t)]. \qquad (2.17)$$

Finally, by substituting (2.17) to (2.14):

$$\frac{\partial L}{\partial \lambda_i} = \sum_{t=1}^{\nu} z_t h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{v}_h \mid z_t) g(\mathbf{w}_h \mid z_t)}{f(\mathbf{x}_h)} \frac{1}{\Psi_{ii}} (w_{ih} - \mu_i - \lambda_i z_t) \tag{2.18}$$

In the same way the partial derivative of the likelihood respect to μ_i is

$$\frac{\partial L}{\partial \mu_i} = \sum_{t=1}^{\nu} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{v}_h \mid z_t) g(\mathbf{w}_h \mid z_t)}{f(\mathbf{x}_h)} \frac{1}{\Psi_{ii}} (w_{ih} - \mu_i - \lambda_i z_t) \tag{2.19}$$

The partial derivative of the likelihood respect to Ψ_{ii}

$$\begin{aligned} \frac{\partial L}{\partial \Psi_{ii}} &= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \frac{\partial f(\mathbf{x}_h)}{\partial \Psi_{ii}} \\ &= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \sum_{t=1}^{\nu} g(\mathbf{v}_h \mid z_t) h(z_t) \frac{\partial g(\mathbf{w}_h \mid z_t)}{\partial \Psi_{ii}} \end{aligned} \tag{2.20}$$

where,

$$\begin{aligned} \frac{\partial g(\mathbf{w}_h \mid z_t)}{\partial \Psi_{ii}} &= (2\pi)^{-1/2} \Psi_{11}^{-1/2} \exp(-\frac{1}{2\Psi_{11}} (w_{1h} - \mu_1 - \lambda_1 z_t)^2) \times \cdots \\ &\quad \times (2\pi)^{-1/2} \Psi_{rr}^{-1/2} \exp(-\frac{1}{2\Psi_{rr}} (w_{rh} - \mu_r - \lambda_r z_t)^2) \times \\ &\quad \frac{\partial}{\partial \Psi_{ii}} [(2\pi)^{-1/2} \Psi_{ii}^{-1/2} \exp(-\frac{1}{2\Psi_{ii}} (w_{ih} - \mu_i - \lambda_i z_t)^2)] \end{aligned} \tag{2.21}$$

and,

$$\begin{aligned} \frac{\partial}{\partial \Psi_{ii}} [(2\pi)^{-1/2} \Psi_{ii}^{-1/2} \exp(-\frac{1}{2\Psi_{ii}} (w_{ih} - \mu_i - \lambda_i z_t)^2)] &= \\ (2\pi)^{-1/2} [(-1/2) \Psi_{ii}^{-3/2} \exp(-\frac{1}{2\Psi_{ii}} (w_{ih} - \mu_i - \lambda_i z_t)^2) + \Psi_{ii}^{-1/2} \times & \\ \exp(-\frac{1}{2\Psi_{ii}} (w_{ih} - \mu_i - \lambda_i z_t)^2)(w_{ih} - \mu_i - \lambda_i z_t)^2 \frac{1}{2\Psi_{ii}^2}] & \end{aligned} \tag{2.22}$$

By substituting (2.22) into (2.21),

$$\frac{\partial g(\mathbf{w}_h \mid z_t)}{\partial \Psi_{ii}} = g(\mathbf{w}_h \mid z_t)[-\frac{1}{2}\Psi_{ii}^{-1} + \frac{1}{2}\Psi_{ii}^{-2}(w_{ih} - \mu_i - \lambda_i z_t)^2] \tag{2.23}$$

By interchanging the summation and substituting (2.23) to (2.20):

$$\begin{aligned}\frac{\partial L}{\partial \Psi_{ii}} &= \sum_{t=1}^{\nu} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{v}_h \mid z_t) g(\mathbf{w}_h \mid z_t)}{f(\mathbf{x}_h)} \times \\ & [-\frac{1}{2}\Psi_{ii}^{-1} + \frac{1}{2}\Psi_{ii}^{-2}(w_{ih} - \mu_i - \lambda_i z_t)^2]\end{aligned} \tag{2.24}$$

Setting the partial derivatives of the continuous part equal to zero, (2.18, 2.19 and 2.24), we get:

$$\begin{aligned}\frac{\partial L}{\partial \hat{\lambda}_i} = 0 \quad &\Longleftrightarrow \quad \sum_{t=1}^{\nu} z_t h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{v}_h \mid z_t) g(\mathbf{w}_h \mid z_t)}{f(\mathbf{x}_h)} \frac{1}{\hat{\Psi}_{ii}} (w_{ih} - \hat{\mu}_i - \hat{\lambda}_i z_t) = 0 \\ &\Longleftrightarrow \quad \sum_{t=1}^{\nu} \sum_{h=1}^{n} z_t h(z_t \mid \mathbf{x}_h)(w_{ih} - \hat{\mu}_i - \hat{\lambda}_i z_t) = 0 \\ &\Longleftrightarrow \quad \sum_{t=1}^{\nu} z_t [r_{2it} - \hat{\mu}_i N_t - \hat{\lambda}_i z_t N_t] = 0 \\ &\Longleftrightarrow \quad \sum_{t=1}^{\nu} z_t r_{2it} - \hat{\mu}_i \sum_{t=1}^{\nu} z_t N_t - \hat{\lambda}_i \sum_{t=1}^{\nu} z_t^2 N_t = 0\end{aligned} \tag{2.25}$$

and,

$$\begin{aligned}\frac{\partial L}{\partial \hat{\mu}_i} = 0 \quad &\Longleftrightarrow \quad \sum_{t=1}^{\nu} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{v}_h \mid z_t) g(\mathbf{w}_h \mid z_t)}{f(\mathbf{x}_h)} \frac{1}{\hat{\Psi}_{ii}} (w_{ih} - \hat{\mu}_i - \hat{\lambda}_i z_t) = 0 \\ &\Longleftrightarrow \quad \sum_{t=1}^{\nu} \sum_{h=1}^{n} h(z_t \mid \mathbf{x}_h)(w_{ih} - \hat{\mu}_i - \hat{\lambda}_i z_t) = 0 \\ &\Longleftrightarrow \quad \sum_{t=1}^{\nu} [r_{2it} - \hat{\mu}_i N_t - \hat{\lambda}_i z_t N_t] = 0 \\ &\Longleftrightarrow \quad \sum_{t=1}^{\nu} r_{2it} - \hat{\mu}_i \sum_{t=1}^{\nu} N_t - \hat{\lambda}_i \sum_{t=1}^{\nu} z_t N_t = 0\end{aligned} \tag{2.26}$$

By solving the system with the equations (2.25) and (2.26) we get explicit formulae for the estimation of the unknown parameters λ_i and μ_i, which are:

$$\hat{\lambda}_i = \frac{(\sum_{t=1}^{\nu} N_t)(\sum_{t=1}^{\nu} z_t r_{2it}) - (\sum_{t=1}^{\nu} r_{2it})(\sum_{t=1}^{\nu} z_t N_t)}{(\sum_{t=1}^{\nu} N_t)(\sum_{t=1}^{\nu} z_t^2 N_t) - (\sum_{t=1}^{\nu} z_t N_t)^2} \tag{2.27}$$

$$\hat{\mu}_i = \frac{(\sum_{t=1}^{\nu} r_{2it})(\sum_{t=1}^{\nu} z_t^2 N_t) - (\sum_{t=1}^{\nu} z_t N_t)(\sum_{t=1}^{\nu} z_t r_{2it})}{(\sum_{t=1}^{\nu} N_t)(\sum_{t=1}^{\nu} z_t^2 N_t) - (\sum_{t=1}^{\nu} z_t N_t)^2} \tag{2.28}$$

and,

$$\begin{aligned}
\frac{\partial L}{\partial \hat{\Psi}_{ii}} = 0 \quad &\Longleftrightarrow \quad \sum_{t=1}^{\nu}\sum_{h=1}^{n} h(z_t \mid \mathrm{x}_h)[-\frac{1}{2}\hat{\Psi}_{ii}^{-1} + \frac{1}{2}\hat{\Psi}_{ii}^{-2}(w_{ih} - \hat{\mu}_i - \hat{\lambda}_i z_t)^2] = 0 \\
&\Longleftrightarrow \quad \sum_{t=1}^{\nu}\sum_{h=1}^{n} h(z_t \mid \mathrm{x}_h)\hat{\Psi}_{ii}^{-1} + \sum_{t=1}^{\nu}\sum_{h=1}^{n} h(z_t \mid \mathrm{x}_h)\hat{\Psi}_{ii}^{-2}(w_{ih} - \hat{\mu}_i - \hat{\lambda}_i z_t)^2 = 0 \\
&\Longleftrightarrow \quad \hat{\Psi}_{ii} = \frac{1}{\sum_{t=1}^{\nu} N_t}[\sum_{t=1}^{\nu}\sum_{h=1}^{n}[w_{ih}(w_{ih} - 2\hat{\mu}_i - 2\hat{\lambda}_i z_t) + (\hat{\mu}_i + \hat{\lambda}_i z_t)^2]h(z_t \mid \mathrm{x}_h)] \\
&\Longleftrightarrow \quad \hat{\Psi}_{ii} = \frac{1}{\sum_{t=1}^{\nu} N_t}\{\sum_{t=1}^{\nu}[\sum_{h=1}^{n} w_{ih}^2 h(z_t \mid \mathrm{x}_h) - 2\hat{\mu}_i \sum_{h=1}^{n} w_{ih} h(z_t \mid \mathrm{x}_h) - \\
&\qquad 2\hat{\lambda}_i z_t \sum_{h=1}^{n} w_{ih} h(z_t \mid \mathrm{x}_h) + (\hat{\mu}_i + \hat{\lambda}_i z_t)^2 \sum_{h=1}^{n} h(z_t \mid \mathrm{x}_h)]\} \\
&\Longleftrightarrow \quad \hat{\Psi}_{ii} = \frac{1}{\sum_{t=1}^{\nu} N_t}\{\sum_{t=1}^{\nu}[r_{3it} - 2\hat{\mu}_i r_{2it} - 2\hat{\lambda}_i z_t r_{2it} + (\hat{\mu}_i + \hat{\lambda}_i z_t)^2 N_t]\} \qquad (2.29)
\end{aligned}$$

Where,

$$r_{2it} = \sum_{h=1}^{n} w_{ih} h(z_t \mid \mathrm{x}_h) \tag{2.30}$$

$$r_{3it} = \sum_{h=1}^{n} w_{ih}^2 h(z_t \mid \mathrm{x}_h) \tag{2.31}$$

$$N_t = \sum_{h=1}^{n} h(z_t \mid \mathrm{x}_h)$$

The equations (2.27) and (2.28) can be written in another form

$$\begin{aligned}
\hat{\lambda}_i &= \frac{(\sum_{t=1}^{\nu} N_t)(\sum_{t=1}^{\nu} z_t r_{2it}) - (\sum_{t=1}^{\nu} r_{2it})(\sum_{t=1}^{\nu} z_t N_t)}{(\sum_{t=1}^{\nu} N_t)(\sum_{t=1}^{\nu} {z_t}^2 N_t) - (\sum_{t=1}^{\nu} z_t N_t)^2} \\
&= \frac{\sum_{t=1}^{\nu}(z_t - \bar{z}.)(\theta_{it} - \bar{\theta}_i.)N_t}{\sum_{t=1}^{\nu}(z_t - \bar{z}.)^2 N_t} \qquad (2.32)
\end{aligned}$$

$$
\begin{aligned}
\hat{\mu}_i &= \frac{(\sum_{t=1}^{\nu} r_{2it})(\sum_{t=1}^{\nu} z_t{}^2 N_t) - (\sum_{t=1}^{\nu} z_t N_t)(\sum_{t=1}^{\nu} z_t r_{2it})}{(\sum_{t=1}^{\nu} N_t)(\sum_{t=1}^{\nu} z_t{}^2 N_t) - (\sum_{t=1}^{\nu} z_t N_t)^2} \\
&= \bar{\theta}_{i\cdot} - \hat{\lambda}_i \bar{z}. \qquad (2.33)
\end{aligned}
$$

where,

$$\theta_{it} = \frac{r_{2it}}{N_t}$$

$$\bar{\theta}_{i\cdot} = \frac{\sum_{t=1}^{\nu} \theta_{it} N_t}{\sum_{t=1}^{\nu} N_t}$$

$$\bar{z}_{\cdot} = \frac{\sum_{t=1}^{\nu} z_t N_t}{\sum_{t=1}^{\nu} N_t}$$

The equations (2.32) and (2.33) could be interpreted as least squares estimates of the regression of the dependent variable θ on the variable z, where the variables z and θ are multiplied by a weight factor N_t and N_t is the number of observations. The regression model is written as $\theta_{it} = \mu_i + \lambda_i z_t$, where $\theta_{it} = \frac{r_{2it}}{N_t}$.

2.2.2 More than one factor latent trait model with mixed manifest variables

The extension of the theory described above to more than one factor latent trait model does not have any theoretical difficulty. However, the full maximum likelihood method is computationally time consuming when more than two factors are fitted.

If there is more than one latent variable the above formulae require some modification. The maximum likelihood equations for the q-latent trait model are given.

The joint distribution of the manifest variables is given by

$$f(\mathbf{x}_h) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{w}_h \mid \mathbf{z}) g(\mathbf{v}_h \mid \mathbf{z}) h(\mathbf{z}) d\mathbf{z} \qquad (2.34)$$

where the z's are assumed to be independent standard normal variables.

The response function for the binary items takes the form:

$$\text{logit}\pi_i(\mathbf{z}) = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j$$

Again here equation (2.34) can be approximated to any practical degree of accuracy by Gauss-Hermite quadrature:

$$f(\mathbf{x}_h) = \sum_{t_1=1}^{\nu_1} \cdots \sum_{t_q=1}^{\nu_q} g(\mathbf{w}_h \mid z_{1t_1}, \cdots, z_{qt_q}) g(\mathbf{v}_h \mid z_{1t_1}, \cdots, z_{qt_q}) h(z_{1t_1}) \cdots h(z_{qt_q})$$

for $h = 1, \cdots, n$

where $z_{1t_1}, \cdots, z_{qt_q}$ are tabled quadrature nodes and $h(z_{1t_1}), \cdots, h(z_{qt_q})$ are the corresponding weights (Stroud and Secrest 1966).

In order to estimate the unknown parameters for the discrete and the continuous part we will maximize the log-likelihood function of the joint distribution of the manifest variables. The maximization procedure is based on the E-M algorithm discussed in section 2.2.2.

The log-likelihood for a random sample of size n will be

$$L = \sum_{h=1}^{n} \log f(\mathbf{x}_h)$$

or

$$L = \sum_{h=1}^{n} \log \sum_{t_1=1}^{\nu_1} \cdots \sum_{t_q=1}^{\nu_q} g(\mathbf{w}_h \mid z_{1t_1}, \cdots, z_{qt_q}) g(\mathbf{v}_h \mid z_{1t_1}, \cdots, z_{qt_q}) h(z_{1t_1}) \cdots h(z_{qt_q})$$

The partial derivatives for the discrete part are

$$\frac{\partial L}{\partial \hat{\alpha}_{i0}} = \sum_{t_1=1}^{\nu_1} \cdots \sum_{t_q=1}^{\nu_q} (r(1, i, t_1, \cdots, t_q) - N(t_1, \cdots, t_q) \pi_i(z_{1t_1}, \cdots, z_{qt_q})) = 0 \tag{2.35}$$

$$\frac{\partial L}{\partial \hat{\alpha}_{ij}} = \sum_{t_1=1}^{\nu_1} \cdots \sum_{t_q=1}^{\nu_q} z_{jt_j}(r(1,i,t_1,\cdots,t_q) - N(t_1,\cdots,t_q)\pi_i(z_{1t_1},\cdots,z_{qt_q})) = 0 \tag{2.36}$$

for $j = 1, \cdots, q$

where,

$$\begin{aligned} r(1,i,t_1,\cdots,t_q) &= h(z_{1t_1})\cdots h(z_{qt_q})\sum_{h=1}^{n} v_{ih} g(\mathbf{w}_h \mid z_{1t_1},\cdots,z_{qt_q}) g(\mathbf{v}_h \mid z_{1t_1},\cdots,z_{qt_q})/f(\mathbf{x}_h) \\ &= \sum_{h=1}^{n} v_{ih} h(z_{1t_1},\cdots,z_{qt_q} \mid \mathbf{x}_h) \end{aligned} \tag{2.37}$$

$$\begin{aligned} N(t_1,\cdots,t_q) &= h(z_{1t_1})\cdots h(z_{qt_q})\sum_{h=1}^{n} g(\mathbf{w}_h \mid z_{1t_1},\cdots,z_{qt_q}) g(\mathbf{v}_h \mid z_{1t_1},\cdots,z_{qt_q})/f(\mathbf{x}_h) \\ &= \sum_{h=1}^{n} h(z_{1t_1},\cdots,z_{qt_q} \mid \mathbf{x}_h) \end{aligned} \tag{2.38}$$

The interpretation of equations (2.37) and (2.38) are equivalent to (2.12) and (2.13) for the one factor latent trait model.

We continue computing the partial derivatives for the parameters of the continuous part.

$$\frac{\partial L}{\partial \hat{\mu}_i} = \sum_{t_1=1}^{\nu_1} \cdots \sum_{t_q=1}^{\nu_q} [r(2,i,t_1,\cdots,t_q) - \hat{\mu}_i N(t_1,\cdots,t_q) - \sum_{j=1}^{q} \hat{\lambda}_{ij} z_{jt_j} N(t_1,\cdots,t_q)] = 0 \tag{2.39}$$

$$\frac{\partial L}{\partial \hat{\lambda}_{ij}} = \sum_{t_1=1}^{\nu_1} \cdots \sum_{t_q=1}^{\nu_q} z_{jt_j}[r(2,i,t_1,\cdots,t_q) - \hat{\mu}_i N(t_1,\cdots,t_q) - \sum_{j=1}^{q} \hat{\lambda}_{ij} z_{jt_j} N(t_1,\cdots,t_q)] = 0 \tag{2.40}$$

for $j = 1, \cdots, q$

By solving the system with the equations (2.39) and (2.40) we get the maximum likelihood estimates for the parameters μ_i and λ_{ij}.

The partial derivative for Ψ_{ii} gives an explicit solution which is

$$\begin{aligned}\hat{\Psi}_{ii} &= \frac{1}{\sum_{t_1=1}^{\nu_1}\cdots\sum_{t_q=1}^{\nu_q}} \times \sum_{t_1=1}^{\nu_1}\cdots\sum_{t_q=1}^{\nu_q}\{r(3,i,t_1,\cdots,t_q) - 2\hat{\mu}_i r(2,i,t_1,\cdots,t_q) \\ &\quad -2\sum_{j=1}^{q}\hat{\lambda}_{ij}z_{jt_j}r(2,i,t_1,\cdots,t_q) - +(\hat{\mu}_i + \sum_{j=1}^{q}\hat{\lambda}_{ij}z_{jt_j})^2 N(t_1,\cdots,t_q)\}\end{aligned} \tag{2.41}$$

where,

$$r(2,i,t_1,\cdots,t_q) = \sum_{h=1}^{n} w_{ih}h(z_{1t_1},\cdots,z_{qt_q} \mid \mathbf{x}_h) \tag{2.42}$$

$$r(3,i,t_1,\cdots,t_q) = \sum_{h=1}^{n} w_{ih}^2 h(z_{1t_1},\cdots,z_{qt_q} \mid \mathbf{x}_h) \tag{2.43}$$

$$N(t_1,\cdots,t_q) = \sum_{h=1}^{n} h(z_{1t_1},\cdots,z_{qt_q} \mid \mathbf{x}_h).$$

2.2.3 Estimation of the parameters

An E-M algorithm is used to obtain the maximum likelihood estimates. This algorithm is iterative and consists of an E step (expectation) followed by a M step (maximization). Dempster, Laird, and Rubin (1977) give the theoretical background of the E-M algorithm. They prove that each iteration of the E-M algorithm not only increases the likelihood, but also that if an instance of the algorithm converges, it converges to a (local) maximum of the likelihood.

The E-M algorithm which presented here is an extension of Bartholomew's modified algorithm presented in Chapter 1 (section 1.3.2) for the case of mixed manifest variables. In the E-M algorithm, described in Bartholomew's book (1987) as a variation of the E-M algorithm, the set of values of the latent variable z which can occur is fixed and we have to predict how many individuals are located at each z.

A software program called LATENT [1] (Moustaki 1995b) has been written in FORTRAN 77 which is based on the program TWOMISS (Albanese and Knott

[1] A brief documentation and description of the program LATENT is given in Appendix A

1992) for handling latent models with binary manifest variables, and which implements the above theory for mixed manifest variables and gives estimates and standard errors for the parameters of interest, α_{i0}, α_{ij}, μ_i, λ_{ij}, and Ψ_{ii}.

2.2.4 E-M algorithm

We define the E-M algorithm as follows

Step 1 Choose starting values for the α_{i0}, α_{ij}, μ_i, λ_{ij} and Ψ_{ii}.

Step 2 Compute the values $r(1, i, t_1, \cdots, t_q)$, $r(2, i, t_1, \cdots, t_q)$, $r(3, i, t_1, \cdots, t_q)$ and $N(t_1, \cdots, t_q)$.

Step 3 Obtain improved estimates of the α_{i0}, α_{ij}, μ_i, λ_{ij} and Ψ_{ii} by solving the equations 2.35, 2.36, 2.39, 2.40, and 2.41 for each item, treating $r(1, i, t_1, \cdots, t_q)$, $r(2, i, t_1, \cdots, t_q)$, $r(3, i, t_1, \cdots, t_q)$ and $N(t_1, \cdots, t_q)$ as given numbers.

Step 4 Return to Step 2 and continue until convergence is attained.

Different initial values can be tried for the parameters of the discrete and the continuous part.

Since the marginal distribution of each row of the manifest variable w is normal with mean μ and covariance matrix Σ, the ML estimate of μ is the mean of the manifest variable w, $(\bar{w})$. So we use the mean of each manifest variable (item) as the initial value of the parameter μ. Starting values for the last two parameters (λ_{ij} and Ψ_{ii}) can be given ad hoc. Rubin and Thayer (1982) in their study used ad hoc and initial values based on PCA.

For the parameters of the continuous part we have derived explicit estimating equations, so we can easily obtain improved estimates as required for Step 3.

Estimating equations for the parameters of the discrete part are obtained by setting the equations (2.35) and (2.36) equal to zero and for each variable i there is a $q + 1$ non-linear equations which can be solved for α_{i0} and α_{ij}. The solution to these equations when the location of each individual on the latent dimension is known is just a logit regression analysis problem. In the programs TWOMISS and LATENT the solution of the non-linear equations is done by a Newton-Raphson procedure, for more details see (Collett (1991), appendix B).

As a criterion for the convergence of the E-M algorithm we compare the relative change in the loglikelihood after each iteration with a very small number, (i.e. 0.0000001). Advantages and drawbacks of the E-M algorithm have been discussed in chapter 1 (section 1.3.2).

When $q > 1$ there is no unique solution because of the fact that orthogonal transformations of the loadings, (α_{ij} and λ_{ij}), leave the value of the joint distribution of the manifest variables unchanged. More specifically for the binary items by premultiplying the loadings α_{ij} by an orthogonal matrix $M_{q\times q}$ we get the transformed values α^*_{ij}. The logit model is then written: $\text{logit}(\pi_i) = (A^* M^{-1})_i \mathbf{z} = A^* \mathbf{z}^*$, where A_i denotes the ith row of the A matrix, ($A_{s\times q} = \{\alpha_{ij}\}$). It appears that the joint distribution of the binary variables will remain unchanged after this orthogonal transformation if the joint distribution of $\mathbf{z}$ and $\mathbf{z}^*$ are the same. It was shown by Lancaster (1954) that if both $\mathbf{z}$ and $\mathbf{z}^*$ are to be independent under orthogonal transformation they must be normal.

Now for the continuous items if we premultiply the loadings λ_{ij} by an orthogonal matrix M we get the transformed λ^*_{ij}. Then the part of the linear factor model which is going to be influenced is again: $\sum_{j=1}^{q} \lambda_{ij} z_j$. After the orthogonal rotation of the loadings that becomes: $(\Lambda^* M^{-1})_i \mathbf{z} = \Lambda^*_i \mathbf{z}^*$, which is the same as in the binary case.

If we put these two results together then simultaneous orthogonal transformations, (rotation), of the coefficients of the mixed model (α_{ij} and λ_{ij}) leave the value of the likelihood unchanged and so they are allowed to be used for finding simple structures in the factor loadings.

2.2.5 Interpretation of the parameters

The parameters α_{i0} and α_{ij} of the discrete part and μ_i, λ_{ij} and Ψ_{ii} of the continuous part are not directly comparable. That is a problem when we come to identify the factors by looking at the factor loadings.

The problem is solved by standardizing the coefficients of the latent variables α_{ij} and λ_{ij} in order to express correlation coefficients between the manifest variable w_i and the latent variable z_j.

Let consider first the parameters of the continuous part. The λ_{ij} denotes covariance between the manifest variable w_i and the latent variable z_j. By dividing λ_{ij} by the square root of the variance of the continuous variable w_i we obtain the correlation between the variable w_i and z_j i.e.:

$$\lambda_{ij}^* = \frac{\lambda_{ij}}{\sqrt{\sum_{j=1}^{q} \lambda_{ij}^2 + \Psi_{ii}}} \tag{2.44}$$

Now for the binary items, let consider the underlying variable model:

$$v_i = \begin{cases} 1 & \text{if } v_i^* \geq \tau_i \\ 0 & \text{if } v_i^* < \tau_i \end{cases}$$

where τ_i are called threshold parameters and

$$v_i^* = \mu_i + \sum_{j=1}^{q} \lambda_{ij} z_j + e_i$$

where, λ_{ij} denotes the covariance between the underlying variable v_i^* and the latent variable z_j. From the equivalence of the response function and the underlying variable approach for binary items, (see Bartholomew 1987, page:104), we get that:

$$\lambda_{ij} = \alpha_{ij} \Psi_{ii}^{1/2}$$

so,

$$\begin{aligned} \text{corr}(v_i^*, z_j) &= \lambda_{ij}^* \\ &= \frac{\text{cov}(v_i^*, z_j)}{\sqrt{\text{var}(v_i^*)}} \\ &= \frac{\alpha_{ij} \Psi_{ii}^{1/2}}{\sqrt{\sum_{j=1}^{q} \alpha_{ij}^2 \Psi_{ii} + \Psi_{ii}}} \\ &= \frac{\alpha_{ij}}{\sqrt{\sum_{j=1}^{q} \alpha_{ij}^2 + 1}} \\ &= \alpha_{ij}^* \end{aligned} \tag{2.45}$$

This reparameterization of the α_{ij} coefficients express the correlation between the underlying variable v_i^* and the latent variable z_j.

The coefficients α^*_{ij} and λ^*_{ij} given from equations (2.45) and (2.44) respectively can be used for giving a unified interpretation of the factor loadings. The standardization of the parameters bring the interpretation close to factor analysis.

For the binary items, Albanese (1990) suggested that when the values of the α_{ij} are greater than 2.5 the response function has a threshold at $z = -\alpha_{i0}/\alpha_{ij}$ and it will be preferred to reparameterize these coefficients by using the formulae given in (2.45). This reparametrization of the discrimination parameters give useful results, in the sense that it showed better behaviour of the likelihood function.

2.2.6 Sampling properties of the maximum likelihood estimates

The E-M algorithm does not yield standard errors of the estimated parameters. From the first order asymptotic theory the maximum likelihood estimates have a sampling distribution which is asymptotically normal. Asymptotically the sampling variances and covariances of the maximum likelihood estimates of the parameters α_{i0} and α_{ij} of the discrete and μ_i, λ_{ij} and Ψ_{ii} of the continuous part are given by the elements of the inverse of the information matrix at the maximum likelihood solution.

In the program LATENT the standard errors of the maximum likelihood estimates are based on an approximation of the above matrix which is given by

$$I(\hat{\beta}) = \{\sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)^2} \frac{\partial f(\mathbf{x}_h)}{\partial \beta_j} \frac{\partial f(\mathbf{x}_h)}{\partial \beta_k}\}^{-1}$$

where β is the vector of the estimated parameters. For more details of this approximation see Chapter 1 (section 1.3.2).

Resampling methods such as bootstrapping or jackknife can be used for calculating standard errors for the estimated parameters but they have not been used in this thesis.

2.2.7 Goodness of fit

A difficult task now is to establish a statistical test for checking the fit of the mixed model. Tests for checking the goodness-of-fit for the binary and the continuous model have already been presented in Chapter 1 (section 1.3.2). None of these statistics can be used directly here.

The goodness-of-fit of the one or two-factor latent trait model has been looked at separately for the discrete and the continuous part. That is, the one- two- and three-way margins of the differences between the observed and expected frequencies under the model are investigated for any large discrepancies for pairs and triples of items which will suggest that the model does not fit well for these combinations of items.

For the continuous part we check the discrepancies between the sample covariance matrix and the one estimated from the model. These two ways for checking the goodness-of-fit of the one- and two-factors model will be used in the chapter with the applications (chapter 4).

Now instead of testing the goodness-of-fit of a specified model we can alternatively use a criterion for selecting among a set of different models. This procedure does not give as any information about the goodness-of-fit for each model but in comparison with other models. For that reason it cannot be considered as a goodness-of-fit measure.

However, a model selection criterion could be used for the determination of the number of factors required. In our case it will be to compare the one factor with the two factor model. Sclove (1987) gives a review of some of the model selection criteria used in multivariate analysis such as the Akaike, Schwarz and Kashap. These criteria take account of the value of the likelihood at the maximum likelihood solution and the number of parameters estimated.

As Sclove (1987) pointed all these criteria take the form:

$$-2\log[maxL(k)] + a(n)m(k) + b(k,n) \tag{2.46}$$

where $L(k)$ is the likelihood of the k_{th} model, n is the sample size, and $m(k)$ is a number of parameters estimated in the k_{th} model. The model with the smallest value of (2.46) compared to the other models is the best one.

Akaike's criterion for the determination of the order of an autoregressive model in time series has been also used for the determination of the number of factors in factor analysis, see Akaike (1987). Akaike's criterion as introduced in Akaike (1969) and (1970) used a final prediction error criterion which in time series models was defined by an estimate of the expected mean square one-step ahead prediction error by the model with parameters estimated with least squares. Akaike (1987) found that in factor analysis the prediction error is the fitted distribution that was evaluated by the likelihood.

The Akaike's criterion come from formula (2.46) for $a(n) = 2$ for all n and $b(k,n) = 0$, i.e.

$$AIC = -2\log[maxL(k)] + 2m(k) \tag{2.47}$$

The Sclove (1987) criterion arises from a Bayesian viewpoint and takes the form:

$$-2\log[maxL(k)] + (\log n)m(k)$$

The Rissanen (1978) criterion takes the form:

$$-2\log[maxL(k)] + \log(\frac{n+2}{24})m(k) + 2\log(k+1)$$

The Kashyap (1982) criterion takes the form:

$$-2\log[maxL(k)] + (\log n)m(k) + \log[detB(k,n)]$$

where $B(k,n)$ is the negative of the matrix of second partial derivatives of the $L(k)$, evaluated at the maximum likelihood solution.

Jöreskog and Sörbom (1993b), refer to other selection criteria such as the CAIC developed by Bozdogan (1987) and the single sample cross-validation index ECVI developed by Cudeck and Browne (1983), which are also functions of the likelihood and the degrees of freedoms. The ECVI criterion requires the split of the sample into two subsamples.

As was pointed by Sclove (1987) and Cudeck and Browne (1983), Akaike's criterion is proven to be more favourable to models with a greater number of parameters, than the Schwarz and the ECVI criterion.

Because of the fact that in the AIC criterion the function of $a(n)$ does not depend on n various researchers consider that it is not a consistent method. However, Sclove (1987) mentions that consistency is an asymptotic property and in reality we only deal with finite sample sizes.

2.2.8 Comments on the model

The limitations of the method presented for handling mixed items within the framework of a latent variable model are:

1. The method as illustrated can be easily extended to fit more than one factor to the set of manifest items but it faces computational problems. For up to two factors the method works satisfactorily

2. There is no statistical test for checking the goodness-of-fit for the overall model. However, Akaike's criterion can be used as a selection model criterion.

The advantages are:

1. The method provides a single analysis for fitting a latent trait model on binary and continuous manifest items, by treating the data as they are.

2. A full maximum likelihood estimation is used for obtaining the parameters of the discrete and the continuous part. The maximum likelihood estimates obtained are consistent and efficient.

3. A unified interpretation of the estimated parameters can be given by standardizing the coefficients in order to express correlation coefficients between the manifest and the latent variables.

4. Our approach in comparison with the underlying variable method which is presented by Muthén, Jöreskog and Sörbom, and Arminger and Küster has the following advantages:

First there is no need in our approach to define an underlying variable for each manifest variable. Muthén's and Jöreskog and Sörbom's approaches use the linear factor model by assuming first that the response is normal and then estimating the correlations required such as tetrachoric and biserial correlation coefficients. The correlation matrix obtained might not be positive definite. Their estimation is a limited information method because it is based on the first- and second-order proportions, whereas our method analyzes the data as they are and so takes into account all the information contained in the data. Their method is limited in the number of items that it can handle because of the large weight matrix needed for estimation of the parameters with the generalized least squares method.

In contrast to Arminger and Küster's approach our method does not require a linear factor model for each underlying variable plus a model for defining measurement relations between the underlying and the manifest variables. Because for the binary items the underlying and the response function approach are equivalent (which is not true for categorical variables) and for the continuous items the measurement relation between the underlying and the manifest variable is the identity the two approaches give equivalent results. However their method has not been implemented by any computer program. MECOSA (Schepers and Arminger 1992) does not incorporate their work.

All the above methods lack the sufficiency properties of our method which derive from the use of models from the exponential family, (sufficiency principle). It was shown in (Bartholomew 1987, page 104) that to every underlying variable model of the above kind there is a corresponding linear model defined in terms of a response function. Now given that the logit model has the sufficiency property and also that the likelihood is simpler in the logit case there is obvious reason for preferring the logit.

In addition factor scores for the individuals in the sample are very easily obtained from the sufficiency principle on which our approach is based as it will be shown in the section below.

2.3 Scaling methods

Scaling methods have been discussed in Chapter 1 (section 1.5) for binary and continuous manifest variables. A modified version of these scaling methods has been used here for the latent variable model with mixed data.

The posterior mean of the latent variable given the whole response pattern (binary and continuous) for the one-factor latent model is the following:

$$E(z \mid \mathrm{x}) = \int_{R_z} z g(\mathrm{x} \mid z) h(z) / f(\mathrm{x}) dz$$

and for the two-factor latent variable model:

$$E(z_1 \mid \mathrm{x}) = \int_{Rz_1} z_1 \int_{Rz_2} g(\mathrm{x} \mid \mathrm{z}) h(\mathrm{z}) / f(\mathrm{x}) dz_2 dz_1$$

$$E(z_2 \mid \mathrm{x}) = \int_{Rz_2} z_2 \int_{Rz_1} g(\mathrm{x} \mid \mathrm{z}) h(\mathrm{z}) / f(\mathrm{x}) dz_1 dz_2$$

The component score is the sum of the component score for the binary part plus the component score for the continuous part, i.e.:

$$\sum_{i=1}^{s} \alpha_{ij} v_i + \sum_{i=1}^{r} \frac{\lambda_{ij}}{\Psi_{ii}} w_i$$

Knott and Albanese (1993) results can be generalized for the latent variable model with mixed items to show that for the one factor model the component score and the posterior mean give the same ranking to individuals.

For the continuous part the conditional distribution of the response pattern $\mathbf{w}$ given z is

$$\begin{aligned} g(\mathbf{w} \mid z) &= \prod_{i=1}^{r} g_i(w_i \mid z) \\ &= \prod_{i=1}^{r} (2\pi)^{-1/2} \Psi_{ii}^{-1/2} \exp\left(-\frac{1}{2\Psi_{ii}}(w_i - \mu_i - \lambda_i z)^2\right) \\ &= \prod_{i=1}^{r} (2\pi)^{-1/2} \Psi_{ii}^{-1/2} \exp\left(-\frac{1}{2\Psi_{ii}}(w_i - \mu_i)^2 + \frac{1}{\Psi_{ii}}(w_i - \mu_i) z \lambda_i - \frac{1}{2}\frac{\lambda_i^2}{\Psi_{ii}} z^2\right) \end{aligned}$$

From this

$$g(\mathbf{0} \mid z) = (2\pi)^{-1/2}\Psi_{ii}^{-1/2} \exp\left(-\frac{1}{2\Psi_{ii}}{\mu_i}^2 - \frac{1}{\Psi_{ii}}\mu_i z\lambda_i - \frac{1}{2\Psi_{ii}}{\lambda_i}^2 z^2\right)$$

Hence,

$$g(\mathbf{w} \mid z) = g(\mathbf{0} \mid z)\prod_{i=1}^{r} \exp\left(-\frac{1}{2\Psi_{ii}}({w_i}^2 - 2w_i\mu_i) + \frac{1}{\Psi_{ii}}w_i z\lambda_i\right) \tag{2.48}$$

For the discrete part the conditional distribution of the manifest variables v given the latent variable z is

$$g(\mathrm{v} \mid z) = \prod_{i=1}^{s} \pi_i(z)^{v_i}(1-\pi_i(z))^{1-v_i}$$

$$g(\mathrm{v} \mid z) = g(\mathbf{0} \mid z)\exp\left(c_o(v) + c_1(v)z\right) \tag{2.49}$$

where, $g(\mathbf{0} \mid z)$ is the probability of a zero response pattern v given the latent variable z, and

$$c_o(v) = \sum_{i=1}^{s} \alpha_{i0}v_i$$

$$c_1(v) = \sum_{i=1}^{s} \alpha_{i1}v_i$$

The joint probability of the manifest variables $x = (w, v)$ may be written as

$$\begin{aligned}
f(\mathrm{x}) &= \int_{-\infty}^{+\infty} g(\mathrm{v} \mid z)g(\mathrm{w} \mid z)h(z)dz \\
&= \int g(\mathbf{0} \mid z)\exp(c_o + c_1 z)g(\mathbf{0} \mid z)\exp\{-\sum_{i=1}^{r}\frac{({w_i}^2 - 2w_i\mu_i)}{2\Psi_{ii}} + \sum_{i=1}^{r}\frac{w_i z\lambda_i}{\Psi_{ii}}\}h(z)dz \\
&= (\exp c_o)\exp(-\frac{1}{2}\sum\frac{({w_i}^2 - 2w_i\mu_i)}{\Psi_{ii}}) * f(\mathbf{0},\mathbf{0})\int h(z \mid \mathbf{0})\exp\{(c_1 + \sum\frac{\lambda_i w_i}{\Psi ii})z\}dz \\
&= \exp(c_o - \frac{1}{2}\sum\frac{({w_i}^2 - 2w_i\mu_i)}{\Psi_{ii}})f(\mathbf{0},\mathbf{0})M_{z|0}(c_1 + \sum\frac{\lambda_i w_i}{\Psi_{ii}}) \qquad (2.50)
\end{aligned}$$

where, $M_{z|0}$ is the moment generating function of the conditional distribution of the latent variable z given a zero response on all items.

From equations (2.48), (2.49), and (2.50) the conditional distribution of z given the response pattern $x = (w, v)$ is:

$$
\begin{aligned}
h(z \mid \mathbf{v}, \mathbf{w}) &= \frac{g(\mathbf{w} \mid z) g(\mathbf{v} \mid z) h(z)}{f(\mathbf{x})} \\
&= \frac{g(\mathbf{0} \mid z) \exp\left(-\frac{1}{2} \sum \frac{(w_i{}^2 - 2w_i\mu_i)}{\Psi_{ii}} + \sum \frac{1}{\Psi_{ii}} w_i z \lambda_i\right) g(\mathbf{0} \mid z) \exp\left(c_o(v) + c_1(v) z\right) h(z)}{\exp\left(c_o - \frac{1}{2} \sum \left(\frac{w_i{}^2 - 2w_i\mu_i}{\Psi_{ii}}\right)\right) f(\mathbf{0}, \mathbf{0}) M_{z|0}(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i)} \\
&= \frac{g(\mathbf{0} \mid z) g(\mathbf{0} \mid z) h(z) \exp(\sum \frac{1}{\Psi_{ii}} \lambda_i w_i z) \exp(c_1 z)}{f(\mathbf{0}, \mathbf{0}) M_{z|0}(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i)} \\
&= \frac{\exp\{(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i) z\} h(z \mid 0)}{M_{z|0}(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i)} \qquad (2.51)
\end{aligned}
$$

From equation (2.51), the moment generating function of the conditional distribution of z given $x = (w, v)$ is

$$
\begin{aligned}
M_{z|x}(t) &= \int_{-\infty}^{+\infty} \exp(tz) h(z \mid \mathbf{x}) dz \\
&= \int \exp(tz) \frac{\exp\{(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i) z\} h(z \mid 0)}{M_{z|0}((c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i)} dz \\
&= \frac{M_{z|0}(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i + t)}{M_{z|0}(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i)} \qquad (2.52)
\end{aligned}
$$

Result 1 If $K_{z|0}(t)$ is the cumulant generating function for the density of z given that all responses are zero, then

$$
\begin{aligned}
E(z \mid \mathbf{x}) &= M'_{z|x}(t) \\
&= K'_{z|x}(t) \\
&= \frac{1}{M_{z|0}(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i)} M'_{z|0}(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i + t) \mid_{t=0} \\
&= K'_{z|0}(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i) \qquad (2.53)
\end{aligned}
$$

and

$$
Var(z \mid \mathbf{x}) = K''_{z|0}(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i) \qquad (2.54)
$$

where the prime and double prime indicate first and second derivatives of the cumulant generating function.

Result 2 $E(z \mid \mathbf{x})$ is a strictly increasing function of $(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i)$, if the variance of the conditional distribution of z given that all responses are zero has variance strictly greater than zero. Knott and Albanese (1993) give this proof for the one logit/probit model for binary data only. From equation (2.53) it follows that $E(z \mid \mathbf{x})$ is a strictly increasing function of the $(c_1 + \sum \frac{\lambda_i}{\Psi_{ii}} w_i)$, if $K'_{z|0}(t)$ is strictly increasing in t. They have shown that using the Cauchy inequality the $K'_{z|0}(t)$ is strictly increasing in t if $K''_{z|0}(t)$ is greater than zero.

There is one more result from the above paper that applies here as well. That is if the conditional distribution of z when all responses are zero is normal, then the conditional distribution of z for any set of responses is normal.

Chapter 3

Latent class model

3.1 Introduction

In this chapter we will discuss the development of a latent class model for mixed manifest variables. The latent class model assumes that the latent space consists of a number of ordered or unordered classes, so the latent space is discrete instead of being continuous as it is for latent trait models. Depending on the level of measurement for the manifest variables we have the latent class model for binary variables and the latent profile model for continuous variables. A systematic investigation of these models was first given in Lazarsfeld and Henry (1968), where they attempted to put latent variable models in a common framework. Bartholomew (1987) as already mentioned in Chapter 1 (section 1.3.2) put the latent variable models in a common framework, using estimation techniques based on maximum likelihood methods.

Many latent class models have been developed over the years. A review of recent theoretical developments and available software for the latent class model are given in Clogg (1993). These developments include analysis of a wide range of manifest variables on different measurement scales, categorical nominal or categorical ordinal latent variables and also reparameterizations of the model that give different insights on the model such as the log-linear or the logit formulation .

In this thesis we will develop a latent class model for mixed manifest variables which are either metric or binary and with categorical-nominal latent variable. Our approach is an extension of Bartholomew's work (1987) for binary and metric variables.

3.2 Latent class model with binary manifest variables

Let $\mathbf{v}$ denote a vector of s binary manifest variables. Let π_{ij} be the probability of a positive response on variable i for an individual in class j, $(i = 1, \cdots, s; j = 0, \cdots, K-1)$ and η_j be the prior probability that a randomly chosen individual is in class j with the constraint that $\sum_{j=0}^{K-1} \eta_j = 1$.

Bartholomew (1987) fitted a latent class model on a number of s binary manifest variables. The marginal distribution of the manifest variables is:

$$f(\mathbf{v}) = \sum_{j=0}^{K-1} \eta_j \prod_{i=1}^{s} \pi_{ij}^{v_i}(1-\pi_{ij})^{1-v_i} \tag{3.1}$$

The log-likelihood for a random sample of size n is:

$$L = \sum_{h=1}^{n} \log\{\sum_{j=0}^{K-1} \eta_j \prod_{i=1}^{s} \pi_{ij}^{v_{ih}}(1-\pi_{ij})^{1-v_{ih}}\} \tag{3.2}$$

Equation (3.2) has to be maximized subject to $\sum_{j=0}^{K-1} \eta_j = 1$, where $\eta_j \geq 0$ and $0 \leq \pi_{ij} \leq 1$.

The maximum likelihood estimates are:

$$\hat{\eta}_j = \sum_{h=1}^{n} h(j \mid \mathbf{v}_h)/n \quad (j = 0, 1, \cdots, K-1) \tag{3.3}$$

and

$$\hat{\pi}_{ij} = \sum_{h=1}^{n} v_{ih} h(j \mid \mathbf{v}_h)/(n\hat{\eta}_j) \quad (i = 1, \cdots, s; j = 0, 1, \cdots, K-1) \tag{3.4}$$

where, $h(j \mid \mathbf{v}_h)$, is the posterior probability that an individual with response pattern $\mathbf{v}_h$ will be allocated to class j given by:

$$h(j \mid \mathbf{v}_h) = \eta_j g(\mathbf{v}_h \mid j)/f(\mathbf{v}_h) \tag{3.5}$$

If the $h(j \mid \mathbf{v}_h)$ were known we could solve equations (3.3) and (3.4) and get the ML estimates for the parameters $\hat{\eta}_j$ and $\hat{\pi}_{ij}$ respectively. The E-M algorithm is used here to derive the ML estimates. The steps of the algorithm are given below:

step 1 Choose initial estimates for the posterior probabilities $h(j \mid \mathbf{v}_h)$

step 2 Use equations (3.3) and (3.4) to obtain a first approximation to $\hat{\eta}_j$ and $\hat{\pi}_{ij}$

step 3 Substitute these estimates to equation (3.5) to obtain improved estimates of $h(j \mid \mathbf{v}_h)$

step 4 Return to step 2 and continue until convergence is attained.

More details about the derivation of the maximum likelihood estimates can be found in (Bartholomew 1987, Chapter 2). Bartholomew also mentions the problem of multiple maxima or local maxima that can be found when fitting latent class models and that the problem increases with the number of classes to be fitted. Aitkin, Anderson, and Hinde (1981) also reported multiple maxima for three or more latent classes fitted on the teaching style data, depending on the different initial values used. Different parameter estimates do not result in a unique interpretation of the classes.

Goodness of fit

Goodness-of-fit for the latent class model can be done by comparing the observed frequencies (O) for each response pattern with the expected frequencies (E) under the fitted model. This comparison is carried out with the chi-square goodness-of-fit statistic given by:

$$X^2 = \sum_{l=1}^{2^s} \frac{(O_l - E_l)^2}{E_l}$$

or the likelihood ratio statistic given by:

$$G^2 = 2\sum_{l=1}^{2^s} O_l \log(\frac{O_l}{E_l})$$

where l denotes the response pattern

or the power-divergence statistic suggested by Read and Cressie (1988), given by:

$$\frac{2}{\lambda(\lambda+1)}\sum_{l=1}^{2^s} O_l\{(\frac{O_l}{E_l})^\lambda - 1\}$$

where l denotes the response pattern and λ is a real-valued parameter chosen by the user. The chi-square statistic and the likelihood ratio test are special cases of the above statistic for values of λ equal to 1 and $\lambda \to 0$ respectively and where 2^s denotes all the possible response patterns.

The above goodness-of-fit statistics are appropriate for use when the number of manifest items to be analyzed, s, is small. When the number of response patterns, 2^s, becomes large there will be cells with very small expected frequencies. It is known that when there is sparseness in the cells of a contingency table, here response patterns, the distribution of the goodness-of-fit statistics presented above are not well approximated by the chi-square distribution. A reference for these types of problem for more general models can be found in Read and Cressie (1988).

A possible way to avoid that problem would be to examine goodness-of-fit for nested models. But Everitt (1988b) and Holt and Macready (1989) checked the distribution of the difference of the likelihood ratio statistic G^2 for nested latent class models and found that it does not have a chi-square distribution.

Some alternative methods for checking the goodness-of-fit of the latent class model have been suggested by Aitkin, Anderson, and Hinde (1981) and Collins, Fidler, Wugalter, and Long (1993). Collins et al. (1993) use a Monte Carlo sampling method in order to find the empirical distribution of the likelihood ratio statistic G^2 instead of assuming that it follows a theoretical distribution, here the χ^2. A number of data sets are generated under the null hypothesis that a latent class model has k classes. For each of the data sets the G^2 value is calculated. These G^2's form an empirical distribution G^2. They applied their procedure to an artificial data set from a four latent class model and they found that the Monte Carlo sampling method worked satisfactorily. However, this method requires a lot of computational time depending on the number of samples generated.

Aitkin et al. (1981) suggested a graphical method which is based on the distribution of the total score $\sum_{i=1}^{s} v_i$ for each individual. If there is only one homogeneous population then the total score will be approximately normal, while if there are K classes and the conditional independence model holds then the distribution of the total score will be approximately normally distributed with K components.

Allocation of individuals into classes

The allocation of individuals into classes is based on the posterior distribution of the latent class given the response pattern of the hth individual, $h(j \mid \mathbf{v}_h)$. An individual with response pattern $\mathbf{v}_h$ will be located to the class with the highest posterior probability compared to the other classes.

The posterior probability for the latent class j is written:

$$\begin{aligned} h(j \mid \mathbf{v}_h) &= \eta_j g(\mathbf{v}_h \mid j)/f(\mathbf{v}_h) \\ &= \frac{\eta_j \prod_{i=1}^{s} \pi_{ij}^{v_{ih}}(1-\pi_{ij})^{(1-v_{ih})}}{\sum_{j=0}^{K-1} \eta_j \prod_{i=1}^{s} \pi_{ij}^{v_{ih}}(1-\pi_{ij})^{(1-v_{ih})}} \end{aligned} \tag{3.6}$$

and the posterior probability for the latent class k is written:

$$\begin{aligned} h(k \mid \mathbf{v}_h) &= \eta_k g(\mathbf{v}_h \mid k)/f(\mathbf{v}_h) \\ &= \frac{\eta_k \prod_{i=1}^{s} \pi_{ik}^{v_{ih}}(1-\pi_{ik})^{(1-v_{ih})}}{\sum_{j=0}^{K-1} \eta_j \prod_{i=1}^{s} \pi_{ij}^{v_{ih}}(1-\pi_{ij})^{(1-v_{ih})}} \end{aligned} \tag{3.7}$$

In order to decide if an individual will be located into class j or class k we look at the ratio:

$$\begin{aligned} \frac{h(j \mid \mathbf{v}_h)}{h(k \mid \mathbf{v}_h)} &= \frac{\eta_j \prod_{i=1}^{s} \pi_{ij}^{v_{ih}}(1-\pi_{ij})^{(1-v_{ih})}}{\eta_k \prod_{i=1}^{s} \pi_{ik}^{v_{ih}}(1-\pi_{ik})^{(1-v_{ih})}} \\ &= \frac{\eta_j}{\eta_k} \exp\{\sum_{i=1}^{s}[\{v_{ih}\log\pi_{ij} + (1-v_{ih})\log(1-\pi_{ij})\} - \\ &\qquad \{v_{ih}\log\pi_{ik} + (1-v_{ih})\log(1-\pi_{ik})\}]\} \end{aligned} \tag{3.8}$$

Furthermore, we allocate an individual into class j if the above ratio is greater than one for all k, i.e.:

$$
\begin{aligned}
&\sum_{i=1}^{s}[v_{ih}\log\pi_{ij} + (1-v_{ih})\log(1-\pi_{ij})] + \log\eta_j > \\
&\quad\sum_{i=1}^{s}[v_{ih}\log\pi_{ik} + (1-v_{ih})\log(1-\pi_{ik})] + \log\eta_k. \qquad (3.9)
\end{aligned}
$$

As we can see the posterior probability depends on the vector $\mathbf{v}$ through a linear function and the allocation of individuals into classes is based on the same linear function, $\sum_{i=1}^{s}[v_{ih}\log\pi_{ij} + (1-v_{ih})\log(1-\pi_{ij})]$. This sort of result holds for all distributions from the exponential family.

3.3 Latent class model with metric manifest variables

A latent class model with metric manifest variables is called a latent profile model. There are two estimation methods for that model, a method based on maximum likelihood and a method based on moments estimators proposed originally by Lazarsfeld and Henry (1968). In the moment estimation method the distribution of the manifest variables is described by specifying the mean and the variance, and for non-normal variables third- or higher-order covariance terms. These moment statistics are used for estimating the unknown parameters of the model. For more details of the estimation see (Lazarsfeld and Henry 1968, Chapter 8). This estimation approach reveals the similarities of the latent profile model to the factor model, (see Bartholomew 1987, Chapter 2, section 2.4).

The maximum likelihood approach for the latent profile model has been described in Bartholomew (1987). This estimation will be extended for the mixed model and for that reason it will be briefly discussed here.

Let $\mathbf{w}$ denotes a vector of r continuous manifest variables. Let μ_{ij} be the mean of the manifest variable i in class j, σ_i^2 be the variance of the manifest variable i assumed constant across classes, $(i = 1, \cdots, r; j = 0, \cdots, K-1)$ and η_j be the prior probability that a randomly chosen individual is in class j with the constraint that $\sum_{j=0}^{K-1}\eta_j = 1$.

The marginal distribution of the manifest continuous variables is:

$$f(\mathbf{w}) = \sum_{j=0}^{K-1} \eta_j \prod_{i=1}^{r} g(w_i \mid j) \tag{3.10}$$

The conditional distribution $g(w_i \mid j)$ was taken, in Bartholomew (1987) to be the normal with mean μ_{ij} and unit variance.

The log-likelihood for a random sample of size n is:

$$L = \sum_{h=1}^{n} \log \sum_{j=0}^{K-1} \eta_j \prod_{i=1}^{r} g(w_{ih} \mid j) \tag{3.11}$$

where, $g(w_{ih} \mid j) \equiv g(w_{ij} \mid \mu_{ij})$

The maximum likelihood estimates are:

$$\hat{\eta}_j = \sum_{h=1}^{n} h(j \mid \mathbf{w}_h)/n \quad (j = 0, 1, \cdots, K-1) \tag{3.12}$$

and

$$\hat{\mu}_{ij} = \sum_{h=1}^{n} w_{ih} h(j \mid \mathbf{w}_h)/(n\hat{\eta}_j) \quad (i = 1, \cdots, s; j = 0, 1, \cdots, K-1) \tag{3.13}$$

where, $h(j \mid \mathbf{w}_h)$, is the posterior probability that an individual with response pattern $\mathbf{w}_h$ will be allocated to class j given by:

$$h(j \mid \mathbf{w}_h) = \eta_j g(\mathbf{w}_h \mid j)/f(\mathbf{w}_h) \tag{3.14}$$

Equations (3.12) and (3.13) can be incorporated into an E-M algorithm to derive the ML estimates. The steps of the E-M algorithm are similar to the ones described for the latent class model for binary manifest variables.

Goodness of fit

To test whether the data arises from a mixture of K normal distributions rather than $K-1$ a likelihood ratio statistic could be used. However, when two models are compared some of the parameters of the submodel are constrained at boundary values of the parameter space under the null hypothesis. As has been discussed by many researchers, (Wolfe 1971, Everitt and Hand 1981, Titterington, Smith, and Makov 1985), in such cases a regularity condition is violated and the likelihood ratio statistic does not follow the chi-square distribution. Simulation results for the form of the distribution of the likelihood ratio statistic can be found in Everitt (1981) and Holt and Macready (1989). Also Aitkin and Rubin (1985) proposed a ML estimation method which places a distribution to the prior probabilities η_j, rather than treating them as parameters and estimate the parameter by maximizing the $\int L(\mu_{ij}, \sigma_i^2, \eta_j \mid \mathbf{w}) d\eta_j$. Their method involves numerical integration and for that reason is computationally disadvantaged compared to the ML method that does not impose a prior distribution on the parameters. The advantage of their method is that when we test for the number of classes under the null hypothesis this is done within the parameter space of the model parameters. However, Quinn, McLachlan, and Hjort (1987) show that even with a prior distribution for the η_j the regularity conditions do not hold.

Allocation into classes

The allocation of individuals into classes, as in the latent class model for binary variables, is based on the posterior probability $h(j \mid \mathbf{w}_h)$, given by equation (3.14).

An individual is more probable to be in class j than k if:

$$\frac{h(j \mid \mathbf{w}_h)}{h(k \mid \mathbf{w}_h)} > 1 \tag{3.15}$$

For the normal model defined in the previous section equation (3.15) becomes:

$$\begin{gathered}
\frac{\eta_j \, (2\pi)^{-\frac{1}{2}r} \exp -\frac{1}{2}\sum_{i=1}^{r}(w_{ih} - \mu_{ij})^2}{\eta_k \, (2\pi)^{-\frac{1}{2}r} \exp -\frac{1}{2}\sum_{i=1}^{r}(w_{ih} - \mu_{ik})^2} > 1 \quad \Longleftrightarrow \\
\sum_{i=1}^{r} w_{ih}\mu_{ij} - \frac{1}{2}\sum_{i=1}^{r} \mu_{ij}^2 + \log \eta_j > \\
\sum_{i=1}^{r} w_{ih}\mu_{ik} - \frac{1}{2}\sum_{i=1}^{r} \mu_{ik}^2 + \log \eta_k
\end{gathered} \tag{3.16}$$

From equation (3.16) we see that the allocation of individuals into classes is also based on a linear function of the manifest variables $\mathbf{w}$.

3.4 Latent class model with mixed manifest variables

Everitt (1988a) and Everitt and Merette (1990) have dealt with the problem of clustering mixed-mode data. They have used both a maximum likelihood method which assumes that the manifest categorical variables are generated by underlying continuous variables and the traditional hierarchical clustering methods such as the complete linkage, group average and Ward's method based on similarities and distance matrices which based on the Euclidean distance calculated from raw data, or from the data standardized to unit variance on each variable, or from the data after each variable has been standardized by its range, and Gower's similarity coefficient (Gower 1971). Using simulation results, (see Everitt and Merette 1990), they found that the hierarchical clustering methods have an unsatisfactory performance compared to the maximum likelihood method. However, the estimation of the model by maximum likelihood requires the evaluation of multidimensional integrals and that restricts the number of categorical variables to one or two.

The model presented by Everitt (1988a) will be described here for reasons of completeness. Suppose there is a vector of $(p = r + s)$ continuous random variables $w_1, \cdots, w_r, w_{r+1}, \cdots, w_p$ with density function:

$$f(\mathrm{w}) = \sum_{j=0}^{K-1} \eta_j MVN_{(r+s)}(\mu_i, \Sigma) \tag{3.17}$$

where k is the assumed number of classes, η_j is the prior probability of each class j or the mixing proportions for which $\sum_{j=0}^{K-1} \eta_j = 1$ and where the $r + s$ variables have a multivariate normal distribution with mean μ_i and covariance matrix Σ taken as constant across classes.

Now suppose that the variables $w_{r+1}, \cdots, w_{r+s}$ are not directly observable, but are related to a set of categorical manifest variables, v, through a threshold model in the following way:

$$v_i = \begin{cases} 1 & \text{if } \alpha_{ij1} < w_{r+i} \leq \alpha_{ij2} \\ 2 & \text{if } \alpha_{ij2} < w_{r+i} \leq \alpha_{ij3} \\ c_i & \text{if } \alpha_{ijc_i} < w_{r+i} \leq \alpha_{ijc_{i+1}} \end{cases}$$

where α_{ijl} are called threshold parameters and these are the ones that generate the manifest categorical variables from the underlying continuous variables, $(w_{r+1}, \cdots, w_{r+s})$ and c_i denote the number of categories of the ith categorical variable and $i = 1, \cdots, s; j = 1, \cdots, K$ and $l = 1, \cdots, c_i$.

The joint density function of the manifest variables is written as:

$$f(\mathbf{w}, \mathbf{v}) = \sum_{j=1}^{K} \eta_j \int_{\alpha_1}^{\beta_1} \cdots \int_{\alpha_s}^{\beta_s} MVN_{(r+s)}(\mu_i, \Sigma) dw_{r+1} \cdots dw_{r+s} \tag{3.18}$$

where $\mathbf{w}' = [w_1, \cdots, w_r]$ and $\mathbf{v}' = [v_1, \cdots, v_s]$

The joint density function of equation (3.18) can be written in an alternative form as:

$$f(\mathbf{w}, \mathbf{v}) = \sum_{j=1}^{K} \eta_j MVN_{(r)}(\mu_i^{(r)}, \Sigma_r) \int_{\alpha_1}^{\beta_1} \cdots \int_{\alpha_s}^{\beta_s} MVN_{(s)}(\mu_i^{(s|r)}, \Sigma_{(s|r)}) dy_1 \cdots dy_s \tag{3.19}$$

The loglikelihood for a random sample of size n is:

$$L = \sum_{h=1}^{n} \log h(\mathbf{w}_h, \mathbf{v}_h) \tag{3.20}$$

For the maximization of the loglikelihood Everitt used several optimization routines such as the Simplex method and a number of quasi-Newton algorithms. The computational difficulties that arise from the evaluation of the integrals limit the number of categorical variables to be analyzed to one or two.

Our formulation of the latent class model does not assume that the binary variables are generated by underlying variables but analyzes the manifest variables as they are. As a result, our method does not involve any numerical integration and that speeds up the estimation procedure and allows a large number of binary and continuous manifest variables to be analyzed.

An extension of the theory presented for the latent trait model, (see Chapter 2), can be used to fit a latent class model on a set of mixed manifest variables. The same notation as before will be used for the manifest variables. In these models we assume that the factor space consists of k classes. That replaces the continuum factor space of a latent trait model. For each class there is an associated probability, η_j. The joint distribution of the manifest variables, using the assumption of conditional independence is a finite mixture of conditional probabilities:

$$f(\mathbf{x}_h) = \sum_{j=0}^{K-1} \eta_j g(\mathbf{w}_h \mid j) g(\mathbf{v}_h \mid j)$$

where $g(\mathbf{w}_h \mid j)$ is the conditional distribution of the vector of manifest continuous variables for the h individual given the class j and $g(\mathbf{v}_h \mid j)$ is the conditional distribution of the vector of manifest binary variables for the h individual given the class j.

Under the assumption of conditional independence and the sufficiency principle, the forms of these conditional distributions are taken from the exponential family and more specifically the conditional distribution of the continuous items, where $g(\mathbf{w}_h \mid j, \mu_{ij}, \sigma_i^2) \equiv g(\mathbf{w}_h \mid j)$, is taken to be:

$$g(\mathbf{w}_h \mid j, \mu_{ij}, \sigma_i^2) = \prod_{i=1}^{r} (2\pi)^{-1/2} \sigma_i^{-1/2} \exp(-\frac{1}{2\sigma_i^2}(w_{ih} - \mu_{ij})^2) \tag{3.21}$$

and the conditional distribution of the binary items is taken to be:

$$g(\mathbf{v}_h \mid j) = \prod_{i=1}^{s} {\pi_{ij}}^{v_{ih}} (1 - \pi_{ij})^{1-v_{ih}} \tag{3.22}$$

where j denotes the class, $(j = 0, \cdots, K-1)$, π_{ij} denotes the probability that an individual who belongs to class j will respond positively to item i, μ_{ij} is the location parameter of the continuous item i in the class j and σ_i^2 is the variance of the i^{th} item taken as constant across classes.

Finally, the log-likelihood for a random sample of size n is written:

$$
\begin{aligned}
L &= \sum_{h=1}^{n} \log f(\mathbf{x}_h) \\
&= \sum_{h=1}^{n} \log \sum_{j=0}^{K-1} \eta_j \prod_{i=1}^{r} [(2\pi)^{-1/2} \sigma_i^{-1/2} \exp(-\frac{1}{2\sigma_i^2}(w_{ih} - \mu_{ij})^2)][\prod_{i=1}^{s} [\pi_{ij}^{v_{ih}} (1 - \pi_{ij})^{1-v_{ih}}]]
\end{aligned}
$$

The above log-likelihood can be maximized using an EM algorithm under the constraint that: $\sum_{j=0}^{K-1} \eta_j = 1$ where $\eta_j \geq 0$. In other words we need to maximize the function:

$$
\Phi = L - \theta(\sum_{j=0}^{K-1} \eta_j - 1) \tag{3.23}
$$

where θ is the Lagrance multiplier.

Finding partial derivatives, we have:

$$
\begin{aligned}
\frac{\partial \Phi}{\partial \eta_j} &= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \frac{\partial f(\mathbf{x}_h)}{\partial \eta_j} - \theta \\
&= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} g(\mathbf{w}_h \mid j) g(\mathbf{v}_h \mid j) - \theta
\end{aligned} \tag{3.24}
$$

$$
\frac{\partial \Phi}{\partial \pi_{ij}} = \sum_{h=1}^{n} \frac{n_j}{f(\mathbf{x}_h)} g(\mathbf{w}_h \mid j) \frac{\partial g(\mathbf{v}_h \mid j)}{\partial \pi_{ij}} \tag{3.25}
$$

where,

$$
\begin{aligned}
\frac{\partial g(\mathbf{v}_h \mid j)}{\partial \pi_{ij}} &= \frac{\partial}{\partial \pi_{ij}} \prod_{i=1}^{s} {\pi_{ij}}^{v_{ih}} \{1 - \pi_{ij}\}^{1-v_{ih}} \\
&= g(\mathbf{v}_h \mid j)[\frac{v_{ih} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})}]
\end{aligned} \tag{3.26}
$$

Hence, by substituting (3.26) into (3.25), we have:

$$
\frac{\partial \Phi}{\partial \pi_{ij}} = \frac{\eta_j}{\pi_{ij}(1 - \pi_{ij})} \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} (v_{ih} - \pi_{ij}) g(\mathbf{w}_h \mid j) g(\mathbf{v}_h \mid j) \tag{3.27}
$$

We carry on by computing the partial derivatives for the parameters of the continuous part.

$$\frac{\partial \Phi}{\partial \mu_{ij}} = \sum_{h=1}^{n} \frac{\eta_j}{f(\mathbf{x}_h)} g(\mathbf{v}_h \mid j) \frac{\partial g(\mathbf{w}_h \mid j)}{\partial \mu_{ij}} \tag{3.28}$$

where,

$$\begin{aligned} \frac{\partial g(\mathbf{w}_h \mid j)}{\partial \mu_{ij}} &= \frac{\partial}{\partial \mu_{ij}} \prod_{i=1}^{r} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(\frac{-1}{2\sigma_i^2}(w_{ih} - \mu_{ij})^2\right) \\ &= \frac{(w_{ih} - \mu_{ij})}{\sigma_i^2} g(\mathbf{w}_h \mid j) \end{aligned} \tag{3.29}$$

by substituting (3.29) into (3.28) we have:

$$\frac{\partial \Phi}{\partial \mu_{ij}} = \frac{\eta_j}{\sigma_i^2} \sum_{h=1}^{n} \frac{(w_{ih} - \mu_{ij})}{f(\mathbf{x}_h)} g(\mathbf{w}_h \mid j) g(\mathbf{v}_h \mid j) \tag{3.30}$$

The partial derivative of the loglikelihood respect to σ_i^2

$$\frac{\partial \Phi}{\partial \sigma_i^2} = \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \sum_{j=0}^{K-1} \eta_j g(\mathbf{v}_h \mid j) \frac{\partial g(\mathbf{w}_h \mid j)}{\partial \sigma_i^2} \tag{3.31}$$

where,

$$\frac{\partial g(\mathbf{w}_h \mid j)}{\partial \sigma_i^2} = g(\mathbf{w}_h \mid j)\left\{\frac{(w_{ih} - \mu_{ij})^2}{2\sigma_i^4} - \frac{1}{2\sigma_i^2}\right\} \tag{3.32}$$

by substituting (3.32) into (3.31) we have

$$\frac{\partial \Phi}{\partial \sigma_i^2} = \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \sum_{j=0}^{K-1} \eta_j \frac{1}{2\sigma_i^2}\left\{\frac{(w_{ih} - \mu_{ij})^2}{\sigma_i^2} - 1\right\} g(\mathbf{v}_h \mid j) g(\mathbf{w}_h \mid j) \tag{3.33}$$

The maximum likelihood equations, (3.24, 3.27, 3.30, 3.33) , can be simplified by expressing them in terms of the posterior distribution $h(j \mid \mathbf{w}_h, \mathbf{v}_h)$. The posterior probability than an individual with response pattern $\mathbf{x}_h = (\mathbf{w}_h, \mathbf{v}_h)$ will be in class j, is given by:

$$h(j \mid \mathbf{w}_h, \mathbf{v}_h) = \eta_j g(\mathbf{w}_h \mid j) g(\mathbf{v}_h \mid j) / f(\mathbf{x}_h) \tag{3.34}$$

Setting the partial derivatives equal to zero, (3.24, 3.27, 3.30, 3.33), and substituting (3.34) in them, we get:

$$\begin{aligned}
\frac{\partial\Phi}{\partial\hat{\eta}_j}=0 \quad &\Longleftrightarrow \quad \sum_{h=1}^{n}\frac{1}{f(\mathbf{x}_h)}g(\mathbf{w}_h \mid j)g(\mathbf{v}_h \mid j)-\theta=0 \\
&\Longleftrightarrow \quad \sum_{h=1}^{n}\frac{h(j \mid \mathbf{w}_h,\mathbf{v}_h)}{\hat{\eta}_j}-\theta=0 \\
&\Longleftrightarrow \quad \sum_{h=1}^{n}h(j \mid \mathbf{w}_h,\mathbf{v}_h)=\theta\hat{\eta}_j \qquad (3.35)
\end{aligned}$$

Summing both sides over j and using $\sum_{j=0}^{K-1}\eta_j=1$ we get that $\theta=n$ and hence equation (3.35) becomes:

$$\hat{\eta}_j=\sum_{h=1}^{n}h(j \mid \mathbf{w}_h,\mathbf{v}_h)/n \quad (j=0,1,\cdots,K-1) \qquad (3.36)$$

Also,

$$\begin{aligned}
\frac{\partial\Phi}{\partial\hat{\pi}_{ij}}=0 \quad &\Longleftrightarrow \quad \frac{\hat{\eta}_j}{\hat{\pi}_{ij}(1-\hat{\pi}_{ij})}\sum_{h=1}^{n}\frac{1}{f(\mathbf{x}_h)}g(\mathbf{w}_h \mid j)g(\mathbf{v}_h \mid j)(v_{ih}-\hat{\pi}_{ij})=0 \\
&\Longleftrightarrow \quad \sum_{h=1}^{n}(v_{ih}-\hat{\pi}_{ij})h(j \mid \mathbf{w}_h\mathbf{v}_h)=0 \\
&\Longleftrightarrow \quad \hat{\pi}_{ij}=\sum_{h=1}^{n}v_{ih}h(j \mid \mathbf{w}_h,\mathbf{v}_h)/(n\hat{\eta}_j) \qquad (3.37)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\Phi}{\partial\hat{\mu}_{ij}}=0 \quad &\Longleftrightarrow \quad \frac{\hat{\eta}_j}{\hat{\sigma}_i^2}\sum_{h=1}^{n}\frac{1}{f(\mathbf{x}_h)}g(\mathbf{w}_h \mid j)g(\mathbf{v}_h \mid j)(w_{ih}-\hat{\mu}_{ij}) \\
&\Longleftrightarrow \quad \sum_{h=1}^{n}(w_{ih}-\hat{\mu}_{ij})h(j \mid \mathbf{w}_h,\mathbf{v}_h)=0 \\
&\Longleftrightarrow \quad \hat{\mu}_{ij}=\sum_{h=1}^{n}w_{ih}h(j \mid \mathbf{w}_h,\mathbf{v}_h)/(n\hat{\eta}_j) \qquad (3.38)
\end{aligned}$$

$$\frac{\partial\Phi}{\partial\hat{\sigma}_i^2}=0 \quad \Longleftrightarrow \quad \sum_{h=1}^{n}\frac{1}{f(\mathbf{x}_h)}\sum_{j=0}^{K-1}\hat{\eta}_j g(\mathbf{v}_h \mid j)g(\mathbf{w}_h \mid j)\frac{1}{2\hat{\sigma}_i^2}\{\frac{(w_{ih}-\hat{\mu}_{ij})^2}{\hat{\sigma}_i^2}-1\}=0$$

$$\Longleftrightarrow \quad \frac{1}{2\hat{\sigma}_i^4}\sum_{h=1}^{n}\sum_{j=0}^{K-1} h(j \mid \mathbf{w}_h, \mathbf{v}_h)(w_{ih} - \hat{\mu}_{ij})^2 - \frac{1}{2\hat{\sigma}_i^2}\sum_{h=1}^{n}\sum_{j=0}^{K-1} h(j \mid \mathbf{w}_h, \mathbf{v}_h) = 0$$

$$\Longleftrightarrow \quad \hat{\sigma}_i^2 = \sum_{h=1}^{n}\sum_{j=0}^{K-1}(w_{ih} - \hat{\mu}_{ij})^2 h(j \mid \mathbf{w}_h, \mathbf{v}_h) / \sum_{h=1}^{n}\sum_{j=0}^{K-1} h(j \mid \mathbf{w}_h, \mathbf{v}_h)$$

$$\Longleftrightarrow \quad \hat{\sigma}_i^2 = \sum_{h=1}^{n}\sum_{j=0}^{K-1}(w_{ih} - \hat{\mu}_{ij})^2 h(j \mid \mathbf{w}_h, \mathbf{v}_h)/n \tag{3.39}$$

3.4.1 EM algorithm

An E-M algorithm is used to obtain the maximum likelihood estimates of the unknown parameters.

If $h(j \mid \mathbf{w}_h, \mathbf{v}_h)$ were known we could solve the ML equations respect to the unknown parameters. Based on that fact, the EM algorithm works as follows:

step 1 Choose initial values for the posterior probabilities $h(j \mid \mathbf{w}_h, \mathbf{v}_h)$.

step 2 Obtain a first approximation for $\hat{\eta}_j$, $\hat{\pi}_{ij}$, $\hat{\mu}_{ij}$ and $\hat{\sigma}_i^2$ from the equations (3.36), (3.37), (3.38) and (3.39).

step 3 Substitute these in (3.34) to obtain a new estimate for $h(j \mid \mathbf{w}_h, \mathbf{v}_h)$.

step 4 Return to step 2 and continue until convergence is attained.

The EM algorithm is considered to have converged when the difference between the value of the loglikelihood in two successive iterations is equal to a very small value, i.e. 0.0000001.

Different initial values for the posterior probability $h(j \mid \mathbf{w}_h, \mathbf{v}_h)$ are used in order to investigate probable multiple or local maximum. The initial allocation of individuals into classes is based on their total score. In order to allow different initial values for the posterior probability we use the total score of each individual based on his responses to the binary items and the total score which is based on the responses to the binary and the metric variables. Hence, the initial value $h(j \mid \mathbf{w}_h, \mathbf{v}_h)$ is 1 for an individual who belongs to class j and 0 otherwise.

A software program called CLASSMIX [1] (Moustaki 1995a) has been written in FORTRAN 77 for fitting a latent class model on a set of mixed manifest variables.

[1]A brief documentation and description of the program is given in the Appendix B

3.4.2 Allocation of individuals into classes

We have already discussed the allocation of individuals into classes for the case where the manifest variables are either binary or metric. In both these cases the allocation was based on the posterior probability. The same is also applied in the latent class model for mixed manifest variables. By combining the results we found above for the binary case, (see equation 3.9) and metric case, (see equation 3.16), we have that an individual will be located to class j and not to k if:

$$\begin{aligned}
\frac{h(j \mid \mathbf{w}, \mathbf{v})}{h(k \mid \mathbf{w}, \mathbf{v})} > 1 \quad &\Longleftrightarrow \\
\frac{\eta_j}{\eta_k} \frac{g(\mathbf{w}_h \mid j) g(\mathbf{v}_h \mid j)}{g(\mathbf{w}_h \mid k) g(\mathbf{v}_h \mid k)} > 1 \quad &\Longleftrightarrow \\
\sum_{i=1}^{r} \{ \frac{w_{ih} \mu_{ij}}{\sigma_i^2} - \frac{1}{2} \frac{\mu_{ij}^2}{\sigma_i^2} \} + \sum_{i=1}^{s} [v_{ih} \log \pi_{ij} + (1 - v_{ih}) \log(1 - \pi_{ij})] + \log \eta_j &> \\
\sum_{i=1}^{r} \{ \frac{w_{ih} \mu_{ik}}{\sigma_i^2} - \frac{1}{2} \frac{\mu_{ik}^2}{\sigma_i^2} \} + \sum_{i=1}^{s} [v_{ih} \log \pi_{ik} + (1 - v_{ih}) \log(1 - \pi_{ik})] + \log \eta_k &
\end{aligned} \tag{3.40}$$

So again here the allocation of individuals is based on a function which is linear on the vector of the manifest variables $(\mathbf{w}, \mathbf{v})$.

3.4.3 Standard errors

As has already been discussed in Chapters 1 and 2 the EM algorithm does not yield standard errors of the estimated parameters. Asymptotically, the sampling variances and covariances of the maximum likelihood estimates of the parameters η_j and π_{ij} of the discrete and μ_{ij}, σ_i^2 of the continuous part are given by the elements of the inverse of the information matrix at the maximum likelihood solution.

The standard errors of the maximum likelihood estimates can be obtained from an approximation of the above matrix which is given by

$$I(\hat{\beta}) = \{ \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)^2} \frac{\partial f(\mathbf{x}_h)}{\partial \beta_j} \frac{\partial f(\mathbf{x}_h)}{\partial \beta_k} \}^{-1}$$

where β is the vector of the estimated parameters.

Bartholomew (1987) found out empirically that this approximation is good for standard errors and less good for covariances.

Resampling methods such as bootstrapping or jackknife can be used for calculating standard errors for the estimated parameters but they have not been used in this thesis.

Chapter 4

Applications

Introduction

The latent trait and the latent class models for mixed observed variables presented in Chapters 2 and 3 respectively have been fitted to four data sets. The analysis presented here is used for illustrating the fit of the mixed model into data sets with different sample sizes and different number of observed variables.

Two of the data sets comprise the responses to five memory questions. The third data set is from the sexual attitudes section of the 1990 British Social Attitudes survey, and the fourth data set is from the environment section of the 1991 British Social Attitudes survey.

In this chapter parameter estimates, scoring methods and measures of goodness of fit will be discussed for the four data sets. The analysis is done with the programs LATENT (Moustaki 1995b) and CLASSMIX (Moustaki 1995a).

The same data sets will be analyzed using the underlying variable approach with the program LISCOMP, (Muthén 1987), in order to allow for comparisons with the results of our approach. We also tried to fit the models using LISREL 8, (Jöreskog and Sörbom 1993a) but it did not give us admissible solutions most of the time and so we decided not to show the results. Neither of these programs LISCOMP and LISREL 8 provide standard errors for the parameter estimates of the models we fitted here.

Lastly, data will be simulated from the estimated model for one of the data sets in order to check the goodness-of-fit of the model to these data.

4.1 Memory questions

The question wording for this data set is given in Appendix C. These are four binary questions that deal with detailed recollection of personal circumstances at the time one hears of an event and one ordinal question on the clarity of the recollection of the event. For this paper the ordinal item is treated as an interval scale variable. The five questions were included by British Market Research Bureau International in their August 1993 face-to-face omnibus survey as part of an LSE Cognitive Laboratory experiment. For 489 individuals the event was the resignation of Thatcher as Prime Minister on November 22, 1990; a different 485 individuals were asked about the disaster at Hillsborough football stadium on April 15, 1989.

These five questions have been analysed in Wright, Gaskell, and O'Muircheartaigh (1994). The objective of the research was to test the assumption that a detailed recollection of one's personal circumstances implies a vivid memory; this is what the theory of "flashbulb memory" postulates. They found that for both events, of the people who said they could remember all four attributes, approximately 55% said their memory was only "fairly clear" or worse. This result does not agree with the hypothesized vividness of "flashbulb memories".

We are interested in the existence of one or more latent variables that could explain the interrelationships among the five items. This can be tested by fitting a latent variable model.

For these five items we fitted a one-factor and a two-factor latent trait model and a two-latent class model.

4.1.1 Thatcher's resignation

First we fit a single latent trait model. The maximum likelihood estimates are given in Table 4.1.

The "discrimination" parameters α_{i1} are large for all the items; and in effect can be considered as threshold functions. Consequently we use the standardized form of

Table 4.1: Thatcher's resignation: Parameter estimates and standard errors for the one-factor latent trait model

Variable v_i	α_{i0}	α_{i1}	π_i	α^*_{i1}
where you were [1]	8.12 (*)	29.7 (*)	0.99	0.99
who you were with [2]	1.62 (0.42)	5.71 (0.95)	0.84	0.98
how you heard about it [3]	3.38 (0.41)	2.57 (0.39)	0.97	0.93
what you were doing [4]	1.13 (0.24)	3.04 (0.38)	0.76	0.95
Variable w_i	μ_i	λ_{i1}	Ψ_{ii}	λ^*_{i1}
vividness of recollection [5]	2.93 (0.05)	0.78 (0.06)	0.75 (0.06)	0.67

* The standard errors estimated are so large as to be untrustworthy.

these coefficients, α^*_{i1}, (see Chapter 2, section 2.2.5). The pattern reveals a general factor. The π_i's show a range of "difficulties" for the four binary items which shows that the median individual has a probability almost 1 of responding positively to items 1 and 3, ($\pi_1 = 0.99, \pi_3 = 0.97$).

The correlation between the observed metric variable w_i and the latent variable z is measured by λ^*_{i1}. The value 0.67 obtained here suggests a strong relationship between the continuous variable and the factor underlying the four binary variables. The value 0.75 is the estimate of the parameter Ψ_{ii}, which is the variance of the error term in the linear factor model, and it is estimated jointly from the continuous and the binary manifest items since there is a single analysis for both types of variables.

We carry on by fitting a two factor latent trait model on the same five items. Table 4.2 gives the maximum likelihood estimates. The standardized alpha coefficients for the first latent variable are all large and positive and the coefficients for the second latent variable contrast items 1 and 2 with 3. The coefficient λ^*_{i1} is equal to 0.57 which is quite large. Figure 4.1 suggests no orthogonal rotation will give a simple and more intuitive interpretation of the variables.

The goodness-of-fit of the model is judged by looking at the one- two- and three-way observed and expected margins of the binary part of the model after the mixed model has been fitted. The discrepancies are measured with the statistic $(O-E)^2/E$. These discrepancies for the two-factor model are very small.

The AIC criterion for the one-factor model is 3017.5 and for the two-factor model

Table 4.2: Thatcher's resignation: Parameter estimates and standard errors for the two-factor latent trait model

Variable v_i	α_{i0}	α_{i1}	α_{i2}	π_i	α_{i1}^*	α_{i2}^*
item [1]	35.4 (*)	136.9 (*)	77.8 (*)	1.00	0.869	0.494
item [2]	1.36 (0.41)	4.15 (1.13)	2.99 (0.82)	0.79	0.796	0.574
item [3]	18.6 (*)	27.0 (*)	-1.48 (1.54)	1.00	0.998	-0.055
item [4]	2.67 (16.2)	6.58 (31.0)	1.12 (0.62)	0.93	0.975	0.166
Variable w_i	μ_i	λ_{i1}	λ_{i2}	Ψ_{ii}	λ_{i1}^*	λ_{i2}^*
item [5]	2.92 (0.05)	0.66 (0.06)	0.41 (0.07)	0.76 (0.06)	0.57	0.36

* The standard errors estimated are so large as to be untrustworthy.

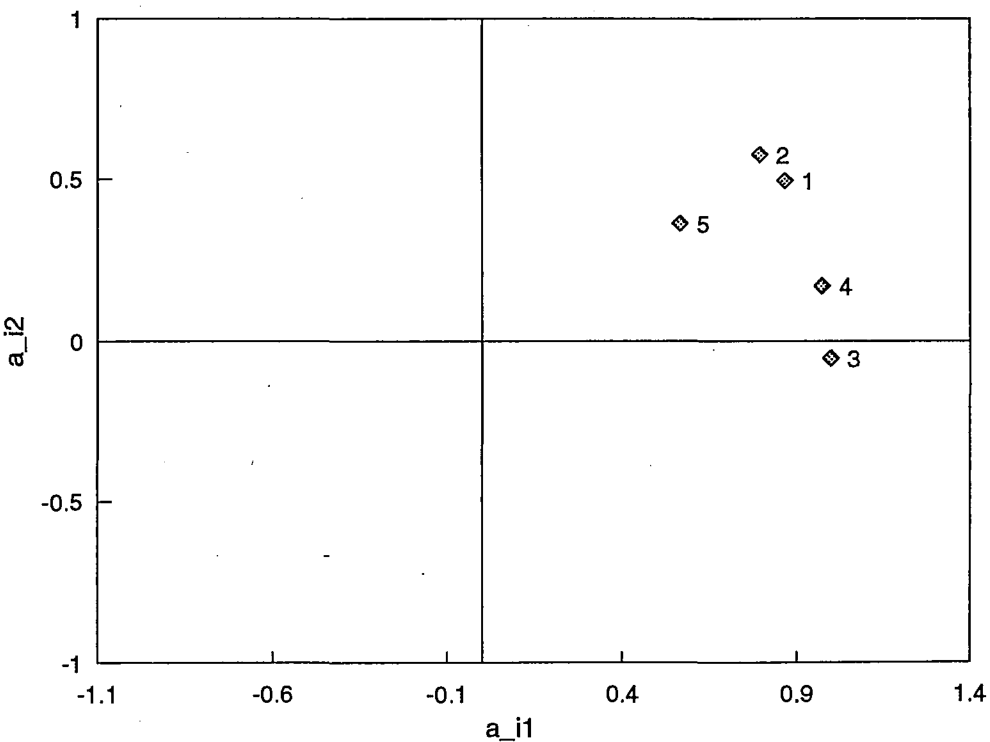

Figure 4.1: Thatcher's resignation - Standardized factor loadings

is 3000.2, suggesting that the two-factor model fits the data better.

Table 4.3 gives the posterior mean for the first and the second latent variable given the response pattern of each individual. The component score is not given here because it depends on the discrimination parameters and since they are very large in that example we think that they will not be particularly meaningful. The ranking of the individuals is based on the posterior mean of the first general factor. The posterior mean takes the same value for some of the response patterns. The reason for that could be the steepness of the response function for the binary items. Figure 4.2 shows that the response patterns with the same value of the posterior mean are located in areas with very small variation. In order to plot the response function of each binary item we set the right linear part of the logit function (i.e. $\alpha_{i0} + \alpha_{i1}z_1 + \alpha_{i2}z_2$) equal to zero. Then for each item the response function can be represented by a line.

A two-latent class model has been fitted to the same data. In this model we assume that the factor space consists of two classes. The parameters π_{ij} for the binary part of the model and μ_{ij} and σ_i^2 for the continuous part are given in Table 4.4. Individuals in class I have very large probabilities of responding positively to the four binary items. Individuals in class II have almost zero probabilities of responding positively to items 1,2 and 4 but still very high probability of responding positively to item 3. The parameters of the continuous item denote that although individuals in class I recollect all their personal circumstances very clearly the estimated mean of this item within class I is only 3.45.

Table 4.3 gives also the allocation of individuals into the two latent classes based on their response patterns. The ranking of individuals based on the latent class model does not perfectly coincide with the ranking based on the latent class model. However the ranking of the latent trait model must be looked at with cautious since many response patterns have the same posterior mean value.

Lastly we want to compare our results with the results obtained from the underlying variable approach. In our case we are not interested in structural equation models (relationships between latent variables) and so only a measurement model is fitted. The estimation method used is a weighted least squares method. The output of the program gives threshold values for the binary variables and the correlation

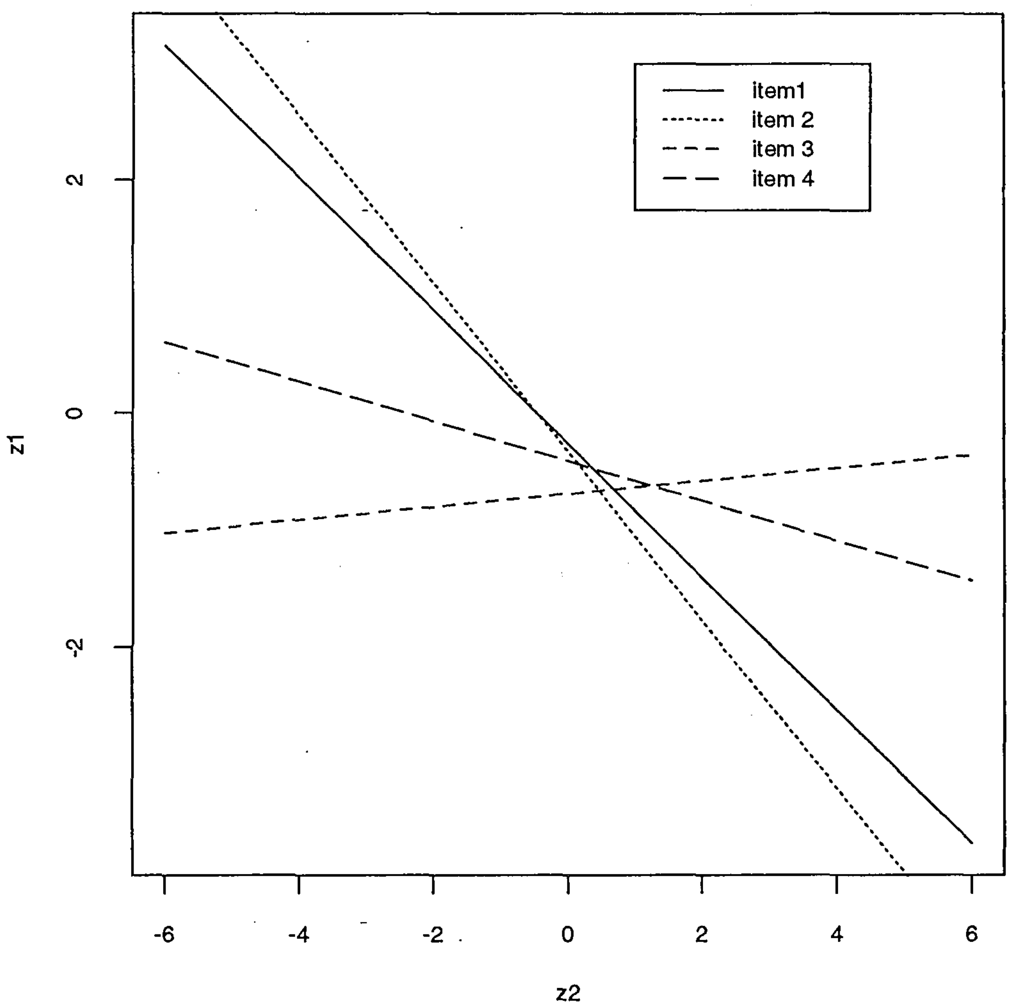

Figure 4.2: Thatcher's resignation: Plot of the response functions of the four binary items

Table 4.3: Thatcher's resignation: Scaling methods

$E(Z_1 \mid \mathbf{x})$	$E(Z_2 \mid \mathbf{x})$	LATENT CLASS	RESPONSE PATTERN
-1.76 (0.38)	-0.37 (0.88)	1	0 0 0 0 1
-1.69 (0.28)	0.00 (0.85)	1	0 0 0 0 2
-1.65 (0.22)	0.35 (0.81)	1	0 0 0 0 3
-1.63 (0.20)	0.66 (0.77)	1	0 0 0 0 4
-1.62 (0.16)	1.40 (0.50)	1	0 1 0 0 2
-1.62 (0.17)	1.48 (0.42)	1	0 1 0 0 3
-0.76 (0.44)	1.67 (0.83)	1	1 1 0 0 4
-0.73 (0.41)	1.82 (0.78)	2	1 1 0 0 5
-0.54 (0.02)	0.61 (0.28)	1	1 0 0 1 2
-0.54 (0.04)	0.54 (0.06)	1	1 0 1 0 1
-0.54 (0.05)	0.54 (0.08)	1	1 0 1 0 2
-0.54 (0.05)	-1.11 (0.67)	1	0 0 1 0 1
-0.54 (0.05)	0.54 (0.11)	1	1 0 1 0 3
-0.54 (0.06)	-0.90 (0.58)	1	0 0 1 0 2
-0.54 (0.11)	-0.44 (0.39)	1	0 1 1 0 2
-0.54 (0.12)	-0.36 (0.46)	1	0 1 1 0 3
-0.54 (0.09)	-0.50 (0.34)	1	0 1 1 0 1
-0.54 (0.10)	-0.59 (0.35)	1	0 0 1 0 5
-0.54 (0.07)	-0.76 (0.48)	1	0 0 1 0 3
-0.54 (0.04)	1.38 (0.66)	2	1 1 0 1 2
-0.54 (0.04)	1.62 (0.68)	2	1 1 0 1 3
-0.54 (0.07)	0.55 (0.14)	2	1 0 1 0 4
-0.53 (0.08)	0.56 (0.18)	2	1 0 1 0 5
-0.53 (0.09)	0.64 (0.34)	2	1 1 1 0 2
-0.53 (0.11)	0.72 (0.43)	2	1 1 1 0 3
-0.52 (0.13)	0.83 (0.52)	2	1 1 1 0 4
-0.51 (0.16)	0.98 (0.60)	2	1 1 1 0 5
-0.06 (0.59)	0.05 (0.68)	2	1 0 1 1 2
-0.04 (0.54)	-1.27 (0.74)	1	0 0 1 1 1
0.04 (0.60)	-0.03 (0.69)	2	1 0 1 1 3
0.05 (0.54)	-1.26 (0.68)	1	0 0 1 1 2
0.14 (0.52)	-1.29 (0.63)	1	0 0 1 1 3
0.15 (0.60)	-0.10 (0.70)	2	1 0 1 1 4
0.21 (0.50)	-1.33 (0.59)	1	0 0 1 1 4
0.22 (0.59)	0.03 (0.73)	2	1 1 1 1 1
0.26 (0.47)	-1.27 (0.68)	2	0 1 1 1 2
0.26 (0.60)	-0.15 (0.71)	2	1 0 1 1 5
0.28 (0.46)	-1.37 (0.56)	1	0 0 1 1 5
0.28 (0.46)	-1.26 (0.72)	2	0 1 1 1 3
0.28 (0.46)	-1.19 (0.84)	2	0 1 1 1 5
0.37 (0.60)	0.08 (0.78)	2	1 1 1 1 2
0.54 (0.63)	0.20 (0.83)	2	1 1 1 1 3
0.79 (0.68)	0.39 (0.88)	2	1 1 1 1 4
1.11 (0.74)	0.64 (0.91)	2	1 1 1 1 5

matrix of the nine items. Here, we only report the values of the factor loadings. The factor loadings of the analysis of the five items with the one factor model are 0.99, 0.97, 0.87, 0.93 and 0.69. The chi-square value obtained is 20.4 with p-value=0.001 indicating a poor fit of the model. However the factor loadings obtained are very close to the results obtained with the program LATENT (see Table 4.1). LISCOMP does not give a solution for the two-factor model. It reports that this is due to a severe Heywood case for variable 3.

4.1.2 Hillsborough football disaster

The same models have been fitted on the second set of memory questions which deal with the Hillsborough football disaster. The maximum likelihood estimates of the one-factor latent trait model are given in Table 4.5.

The π column shows that items 1,2 and 3 have very high probabilities of positive responses from the median individual. We may expect more people to recollect this event than Thatcher's resignation since no other football disasters have happened since then but many political ėvents have occurred. The coefficients α^*_{i1} are all large as we would expect with a general factor. The parameter λ^*_{11} here is also high, $\lambda^*_{11} = 0.53$.

Table 4.6 gives the maximum likelihood estimates of the two-factor latent trait model. For the parameters of the discrete part we see that the loadings of the first factor are all very large and the loadings of the second factor discriminates between item 1 and items 3 and 4. Figure 4.3 plot the standardized factor loadings of the two latent variables. No straight forward rotation is emerged from the plot.

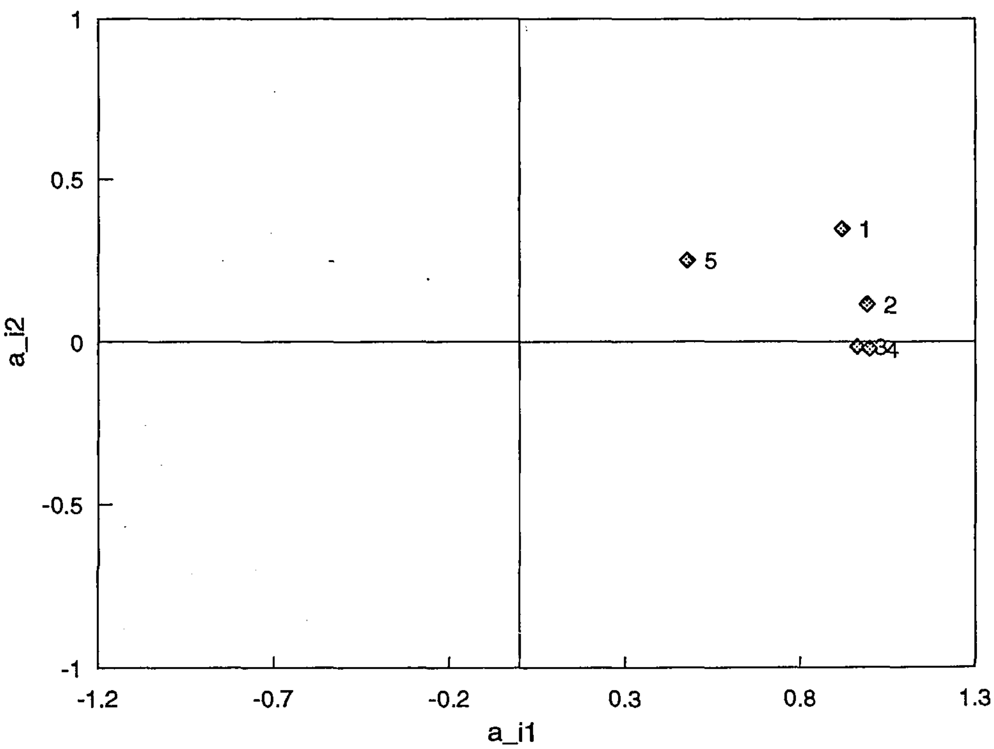

Figure 4.3: Hillsborough disaster - Standardized factor loadings

Table 4.4: Thatcher's resignation: Parameter estimates for the two-latent class model

Variable v_i	$\hat{\pi}_{i1}$	$\hat{\pi}_{i2}$	
where you were [1]	0.979	0.065	
who you were with [2]	0.935	0.056	
how you heard about it [3]	0.979	0.657	
what you were doing [4]	0.892	0.179	
$\hat{\eta}_j$	0.606	0.394	
Variable w_i	$\hat{\mu}_{i1}$	$\hat{\mu}_{i2}$	$\hat{\sigma}_i^2$
vividness of recollection [5]	3.45	2.07	0.911

Table 4.5: Hillsborough disaster: Parameter estimates and standard errors for the one-factor latent trait model

Variable v_i	α_{i0}	α_{i1}	π_i	α^*_{i1}
where you were [1]	4.41 (1.54)	6.45 (2.33)	0.99	0.988
who you were with [2]	2.91 (0.73)	5.29 (1.31)	0.95	0.983
how you heard about it [3]	5.55 (0.83)	3.24 (0.61)	0.99	0.955
what you were doing [4]	1.39 (0.19)	2.03 (0.28)	0.80	0.897
Variable w_i	μ_i	λ_{i1}	Ψ_{ii}	λ^*_{i1}
vividness of recollection [5]	3.19 (0.05)	0.56 (0.05)	0.75 (0.06)	0.53

Table 4.6: Hillsborough disaster: Parameter estimates and standard errors for the two-factor latent trait model

Variable v_i	α_{i0}	α_{i1}	α_{i2}	π_i	α^*_{i1}	α^*_{i2}
item [1]	4.01 (1.55)	5.17 (1.02)	1.95 (2.20)	0.98	0.921	0.347
item [2]	13.0 (*)	23.1 (2.09)	2.68 (*)	1.00	0.992	0.115
item [3]	6.27 (1.40)	3.61 (1.23)	-0.07 (0.93)	0.99	0.964	-0.017
item [4]	10.7 (*)	19.53 (0.32)	-0.43 (*)	1.00	0.998	-0.022
Variable w_i	μ_i	λ_{i1}	λ_{i2}	Ψ_{ii}	λ^*_{i1}	λ^*_{i2}
item [5]	3.21 (0.05)	0.50 (0.11)	0.25(0.06)	0.75 (0.07)	0.48	0.25

* The standard errors estimated are so large as to be untrustworthy.

The AIC criterion for the one-factor model is 2832.1 and for the two-factor model is 2834.8. Here, the one-factor model shows marginally better fit compared to the two-factor model. The discrepancies in the one- two- and three-way margins of the two-factor model show slight improvement on the one-factor model.

From Table 4.7 we see again as with the previous example that there are quite a few response patterns which have exactly the same value of $E(Z_1 \mid \mathbf{x})$ with zero standard errors. Figure 4.4 gives the plot of the response functions of the four binary items. They are threshold functions and we observe that the response patterns with the same value of the posterior mean are located in areas with zero variation.

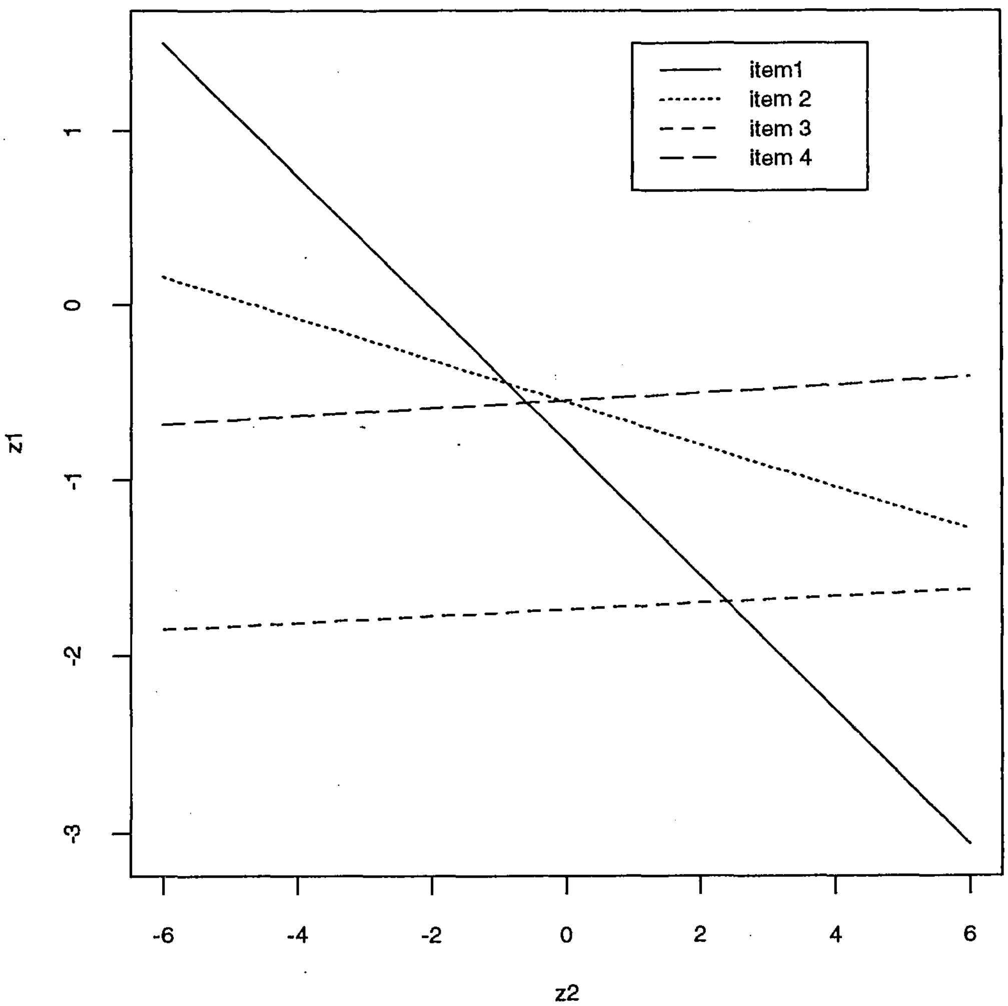

Figure 4.4: Hillsborough disaster: Plot of the response functions of the four binary items

The parameter estimates for the two latent class model are given in Table 4.8. The interpretation of the results is very similar to that for Thatcher's resignation.

Table 4.7: Hillsborough disaster: Scaling methods

$E(Z_1 \mid \mathbf{x})$	$E(Z_2 \mid \mathbf{x})$	CSCORE1	CSCORE2	LATENT CLASS	RESPONSE PATTERNS
-2.00 (0.56)	-0.38 (0.95)	0.66	0.34	1	0 0 0 0 1
-1.84 (0.46)	-0.11 (0.93)	1.32	0.67	1	0 0 0 0 2
-1.74 (0.37)	0.15 (0.92)	1.97	1.01	1	0 0 0 0 3
-1.68 (0.29)	0.41 (0.90)	2.63	1.35	1	0 0 0 0 4
-1.48 (0.38)	1.28 (0.99)	7.14	2.96	1	1 0 0 0 3
-1.44 (0.43)	-0.62 (0.94)	4.27	0.27	1	0 0 1 0 1
-1.35 (0.49)	-0.42 (0.94)	4.93	0.61	1	0 0 1 0 2
-1.26 (0.52)	-0.26 (0.95)	5.59	0.95	1	0 0 1 0 3
-1.16 (0.54)	-0.14 (0.96)	6.24	1.28	1	0 0 1 0 4
-1.08 (0.55)	-0.04 (0.98)	6.90	1.62	1	0 0 1 0 5
-0.66 (0.34)	-0.23 (0.78)	10.09	2.56	1	1 0 1 0 2
-0.64 (0.32)	-0.13 (0.79)	10.75	2.89	1	1 0 1 0 3
-0.63 (0.31)	-0.02 (0.81)	11.41	3.23	1	1 0 1 0 4
-0.54 (0.00)	-0.16 (0.60)	28.12	3.29	1	0 1 1 0 2
-0.54 (0.00)	-0.04 (0.62)	28.77	3.63	1	0 1 1 0 3
-0.54 (0.00)	-1.41 (0.71)	23.83	-0.16	1	0 0 1 1 1
-0.54 (0.00)	0.09 (0.63)	29.43	3.97	1	0 1 1 0 4
-0.54 (0.00)	-1.24 (0.69	24.48	0.18	1	0 0 1 1 2
-0.54 (0.00)	-1.09 (0.66)	25.14	0.52	1	0 0 1 1 3
-0.54 (0.00)	-0.95 (0.63)	25.80	0.86	1	0 0 1 1 4
-0.54 (0.00)	0.63 (0.67)	33.28	5.24	2	1 1 1 0 2
-0.54 (0.00)	-0.53 (0.54)	29.65	2.13	2	1 0 1 1 2
-0.54 (0.00)	0.79 (0.70)	33.94	5.57	2	1 1 1 0 3
-0.54 (0.00)	-0.43 (0.55)	30.31	2.46	2	1 0 1 1 3
-0.54 (0.00)	-0.33 (0.56)	30.96	2.80	2	1 0 1 1 4
-0.54 (0.00)	-0.22 (0.59)	31.62	3.14	2	1 0 1 1 5
-0.54 (0.00)	0.96 (0.73)	34.60	5.91	2	1 1 1 0 4
-0.54 (0.00)	1.15 (0.77)	35.25	6.25	2	1 1 1 0 5
-0.49 (0.22)	0.43 (0.69)	49.22	4.88	2	1 1 0 1 2
-0.47 (0.27)	0.57 (0.71)	49.88	5.21	2	1 1 0 1 3
-0.43 (0.33)	-0.49 (0.80)	47.67	2.86	2	0 1 1 1 2
-0.41 (0.35)	-0.38 (0.81)	48.33	3.20	2	0 1 1 1 3
-0.39 (0.38)	-0.28 (0.83)	48.99	3.54	2	0 1 1 1 4
0.31 (0.60)	-0.20 (0.96)	52.84	4.81	2	1 1 1 1 2
0.49 (0.63)	-0.03 (0.95)	53.49	5.15	2	1 1 1 1 3
0.72 (0.68)	0.18 (0.95)	54.15	5.48	2	1 1 1 1 4
1.02 (0.74)	0.42 (0.96)	54.81	5.82	2	1 1 1 1 5

Individuals in class II have higher probabilities of responding positively to items 1, 3 and 4.

The allocation of individuals in the two classes is given in Table 4.7. There is no disagreement in the ranking of the individuals between the latent class and the latent trait model.

Table 4.8: Hillsborough disaster: Parameter estimates for the two-latent class model

Variable v_i	$\hat{\pi}_{i1}$	$\hat{\pi}_{i2}$	
where you were [1]	0.979	0.141	
who you were with [2]	0.933	0.090	
how you heard about it [3]	0.995	0.774	
what you were doing [4]	0.863	0.274	
$\hat{\eta}_j$	0.724	0.276	
Variable w_i	$\hat{\mu}_{i1}$	$\hat{\mu}_{i2}$	$\hat{\sigma}_i^2$
vividness of recollection [5]	3.44	2.53	0.898

The five items were analyzed with the program LISCOMP. The factor loadings of the one and two-factor models are given in Table 4.9.

Table 4.9: Hillsborough disaster: Parameter estimates from LISCOMP

Variable v_i	one-factor	two-factor	
where you were [1]	0.98	0.84	0.45
who you were with [2]	0.92	0.86	0.43
how you heard about it [3]	0.89	0.62	0.66
what you were doing [4]	0.90	0.40	0.86
vividness of recollection [5]	0.56	0.49	0.27
	$\chi^2 = 23.0$	χ^2=5.42	
	p-value=0.000	p-value=0.02	

The two-factor model indicates a better fit than the one-factor model at 2% significance level. To compare the factor loadings of the two solutions (LISCOMP and LATENT) we orthogonally rotate the two solutions in order to find the best matching factors. The results of the rotation are given in Table 4.10 and they show similar estimated coefficients.

For investigating whether there is a real difference between the loading patterns of the two memory experiments (see Table 4.2 and Table 4.6) we orthogonally rotate the two solutions to find the best matching factors. The rotation results are given in Table 4.11. It appears that there is not much difference in the two solutions, though some differences can be expected since we analyze results from two different samples where individuals were asked the same five memory questions but on different subjects.

4.2 Sexual Attitude Questions, BSA 1990

The third data set is from the sexual attitudes section of the British Social Attitudes, 1990, Survey, (Brook, Taylor, and Prior 1991). The question wording for the nine variables which have been extracted for the analysis are given in Appendix D in the same order that they are going to be analyzed. There were 1121 individuals who were asked questions on sexual relationships.

If we exclude the responses "depends/varies", "don't know" and "not answered" from the above items then items 1 to 6 are binary items with response categories 1 for agree and 0 for disagree and items 7 to 9 are five point scale items with responses "always wrong", "mostly wrong", "sometimes wrong", "rarely wrong" and "not wrong at all". The items 7,8 and 9 will be treated as metric variables.

First we fit a one-factor latent trait model on these items. Parameter estimates are given in Table 4.12. The discrepancies in the one- two- and three-way margins of the observed and expected frequencies of the binary items only, show a very bad fit of the model, especially on the responses that contain items 5 and 6. Also the estimated covariance matrix compared to the sample covariance matrix of the continuous items shows big differences. For response (1,1), the discrepancies for items (5,6) are equal to 56.2. For response (1,0), the discrepancies for items (5,2), (5,3), (5,4), (6,2), (6,3), (6,4) and (6,5) are respectively equal to 15.9, 18.1, 4.2, 25.9, 33.8, 12.1 and 50.5. For response (0,1), the discrepancy for items (6,5) is 22.2. For response (1,1,1), the discrepancies for items (1,5,6), (2,5,6), (3,5,6), and (4,5,6) are respectively 42.8, 30.5, 34.4 and 38.2. Anywhere else the fit was good.

The items on adoption of babies by female and male homosexual couples have a very small probability of a positive response from the median individual, 0.10 and 0.02 respectively. In addition the item on homosexual relations has a smaller mean than the before marriage question and the highest loading among the continuous items. From the π_i column we see that people are more liberal in accepting homosexuals in higher education than in schools.

The bad fit of the model suggests the possibility of an additional dimension and thus the fit of a two-factor model on these items. The results are in Table 4.13.

Table 4.10: Hillsborough disaster: Rotation of the factor loadings obtained from the programs LISCOMP and LATENT

Variable i	LISCOMP α^*_{i1}	solution α^*_{i2}	LATENT α^*_{i1}	solution α^*_{i2}
where you were [1]	0.934	0.203	0.961	0.215
who you were with [2]	0.930	0.233	0.998	-0.025
how you heard about it [3]	0.903	-0.102	0.952	-0.152
what you were doing [4]	0.860	-0.395	0.985	-0.161
vividness of recollection [5]	0.547	0.115	0.510	0.180

Table 4.11: Rotation of the standardized solutions of the two memory experiments

Variable i	Thatcher α^*_{i1}	resignation α^*_{i2}	Hillsborough α^*_{i1}	disaster α^*_{i2}
where you were [1]	0.981	0.194	0.961	0.212
who you were with [2]	0.944	0.273	0.998	-0.028
how you heard about it [3]	0.900	-0.430	0.952	-0.155
what you were doing [4]	0.972	-0.125	0.985	-0.164
vividness of recollection [5]	0.659	0.143	0.511	0.179

Table 4.12: Sexual attitudes: Parameter estimates and standard errors for the one-factor latent trait model

Variable v_i	α_{i0}	α_{i1}	π_i	α^*_{i1}
SEXLAW [1]	1.65 (0.09)	0.55 (0.11)	0.84	0.48
GAYTEAS [2]	-0.57 (0.39)	9.61 (1.30)	0.36	0.99
GAYTEAH [3]	1.75 (0.89)	11.9 (3.55)	0.85	1.00
GAYPUB [4]	0.90 (0.16)	3.72 (0.32)	0.71	0.97
FGAYADP [5]	-2.16 (0.15)	1.68 (0.17)	0.10	0.86
MGAYADP [6]	-3.74 (0.26)	2.38 (0.25)	0.02	0.92
Variable w_i	**μ_i**	**λ_{i1}**	**Ψ_i**	**λ^*_{i1}**
PMS [7]	3.66 (0.05)	0.53 (0.06)	1.87 (0.12)	0.36
EXMS [8]	1.60 (0.03)	0.21 (0.03)	0.66 (0.02)	0.26
SAME SEX [9]	2.02 (0.06)	0.90 (0.06)	1.44 (0.07)	0.60

The two-factor model has improved significantly the fit on the two- and three-way margins of the observed and the expected frequencies of the binary responses and the differences on the sample and estimated covariance matrix of the continuous items. The discrepancies given above for the one-factor model went down to values less than one. The same conclusion is reached by looking at the AIC criterion. For the one-factor model the AIC-criterion is 15972.1 and for the two-factor model the AIC criterion is 15679.5 suggesting that the two-factor model fits the data better than the one-factor model.

The interpretation of the "difficulty" parameters α_{i0} remains the same. The standardized factor loadings, α^*_{ij}, λ^*_{ij}, are recommended for the mixed model because they all somehow express correlations between the latent and the observed variables and so they allow comparisons between the binary and the metric variables. A plot of the standardized factor loadings is given in Figure 4.5. From the plot it emerges that items on adoption of babies by homosexuals (5 and 6) load heavily on the horizontal axis and items on homosexuals teaching and obtaining positions in public life (2,3 and 4) load heavily on the vertical axis. That implies that homosexuality is a two-dimensional issue. Items 7,8 and 9 are somewhere in the middle of these two dimensions. Furthermore items 2,3 and 4 can be considered as measuring the degree of sexual prejudice of individuals and so that dimension could indicate conservatism or liberalism of individuals. On the other hand items 5 and 6 are clearly related with homosexuality and these do not have to do necessarily with the degree of individual's conservatism.

We carried out a latent class analysis to see if we would arrive at similar results. We fitted first a two-latent class model to the items. The parameter estimates are given in Table 4.14. From the estimated $\hat{\pi}_{ij}$ parameters, we see that item 1,7 and 8 do not discriminate much between the two classes. Again the items on adoption and the after marriage relation have quite low probabilities of a positive response in the second class where individuals are more liberal concerning the rest of the items. Anyhow from the $\hat{\eta}_j$ parameter we see that 73% of the individuals in the sample belong to the "conservative" class and only 27% to the "liberal" one.

The fit of the two-latent class model was not good when we look at the observed and expected frequencies under the fitting model for each response pattern of the

Table 4.13: Sexual attitudes: Parameter estimates and standard errors for the two-factor latent trait model

Variable v_i	α_{i0}	α_{i1}	α_{i2}	π_i	α^*_{i1}	α^*_{i2}
SEXLAW [1]	1.67 (0.08)	0.20 (0.11)	0.53 (0.10)	0.84	0.18	0.46
GAYTEAS [2]	-0.73 (0.23)	4.56 (0.28)	8.65 (0.49)	0.32	0.46	0.87
GAYTEAH [3]	1.45 (0.18)	4.67 (0.34)	8.50 (0.44)	0.81	0.48	0.87
GAYPUB [4]	0.89 (0.10)	1.90 (0.17)	2.86 (0.19)	0.71	0.53	0.80
FGAYADP [5]	-4.28 (0.15)	4.94 (0.23)	0.15 (0.17)	0.01	0.98	0.03
MGAYADP [6]	-24.8 (0.36)	16.2 (0.22)	0.33 (0.48)	0.00	1.00	0.02
Variable w_i	μ_i	λ_{i1}	λ_{i2}	Ψ_i	λ^*_{i1}	λ^*_{i2}
PMS [7]	3.67 (0.04)	0.66 (0.06)	0.19 (0.05)	1.69 (0.10)	0.45	0.13
EXMS [8]	1.60 (0.02)	0.26 (0.03)	0.07 (0.03)	0.63 (0.02)	0.31	0.08
SAME SEX [9]	2.03 (0.04)	0.80 (0.05)	0.52 (0.05)	1.35 (0.07)	0.53	0.35

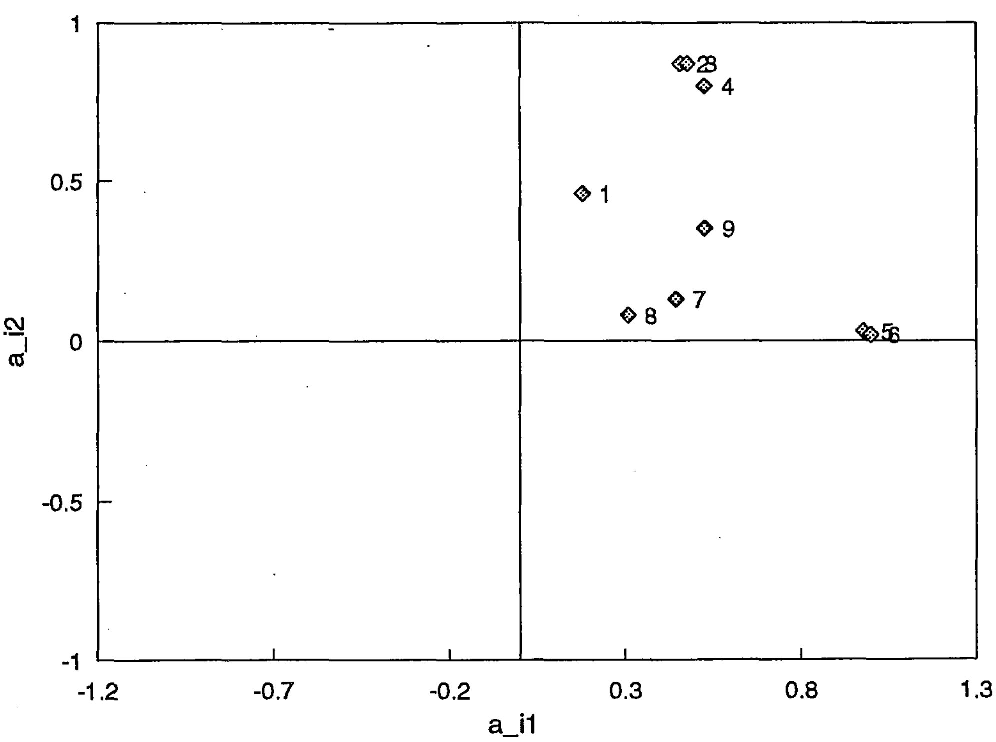

Figure 4.5: Sexual attitudes - Standardized factor loadings

binary items only. A three-latent class model is fitted next. The parameter estimates of the three-latent class model are given in Table (4.15).

Table 4.14: Sexual attitudes: Parameter estimates for the two-latent class model

Variable v_i	$\hat{\pi}_{i1}$	$\hat{\pi}_{i2}$	
SEXLAW [1]	0.786	0.937	
GAYTEASC [2]	0.325	0.865	
GAYTEAH [3]	0.412	0.902	
GAYPUB [4]	0.476	0.877	
FGAYADPT [5]	0.099	0.435	
MGAYADPT [6]	0.027	0.329	
$\hat{\eta}_j$	0.730	0.270	
Variable w_i	$\hat{\mu}_{i1}$	$\hat{\mu}_{i2}$	$\hat{\sigma}_i^2$
PMS [7]	3.374	4.484	1.916
EXMS [8]	1.487	1.909	0.667
SAME SEX [9]	1.203	4.298	0.372

By looking the parameter estimates it looks as if the first class has been split into two classes. The 3-latent class model gives a better fit than the 2-latent class model. Although the third class is considered to be the most "liberal" one again here the items on adoption have relatively small probabilities to receive a positive response from an individual in class 3 compare to the rest of the items. Item 9, "homosexual relations" is the one that discriminates better between class 2 and 3. These results are in agreement with the results we get from the two-latent trait model, in the sense that "homosexuality" appears to be a different issue from "liberalism".

The same items have been analyzed by de Menezes and Bartholomew (1996); they treated all items as binary. Our results are consistent with theirs although our analysis by treating items 7, 8 and 9 as discrete variables avoids the arbitrariness of their dichotomization. In their study they analyzed the data twice once with the middle point of the metric items to be "yes" and once to be "no". Our analysis avoids this work.

The nine observed variables on sexual attitudes were analyzed also with LISCOMP. The results we obtained from the two-factor model are given in Table 4.16.

Since the solutions obtained from both the program LATENT (see Table 4.13)

Table 4.15: Sexual attitudes: Parameter estimates for the three-latent class model

Variable v_i	$\hat{\pi}_{i1}$	$\hat{\pi}_{i2}$	$\hat{\pi}_{i3}$	
SEXLAW [1]	0.762	0.846	0.937	
GAYTEASC [2]	0.007	0.851	0.841	
GAYTEAH [3]	0.056	0.980	0.877	
GAYPUB [4]	0.020	0.915	0.859	
FGAYADPT [5]	0.053	0.164	0.505	
MGAYADPT [6]	0.009	0.065	0.390	
$\hat{\eta}_j$	0.448	0.340	0.212	
Variable w_i	$\hat{\mu}_{i1}$	$\hat{\mu}_{i2}$	$\hat{\mu}_{i3}$	$\hat{\sigma}_i^2$
PMS [7]	3.190	3.720	4.621	1.864
EXMS [8]	1.424	1.638	1.918	0.666
SAME SEX [9]	1.139	1.585	4.665	0.365

Table 4.16: Sexual attitudes: Parameter estimates from LISCOMP

Items	factor 1	factor 2
SEXLAW	0.301	0.138
GAYTEASC	0.926	0.341
GAYTEAH	0.949	0.298
GAYPUB	0.856	0.294
FGAYADPT	0.277	0.869
MGAYADPT	0.286	1.030
PMS	0.205	0.474
EXMS	0.145	0.328
SAME SEX	0.427	0.479

and LISCOMP (see Table 4.16) are arbitrary (no constraints imposed) we orthogonally rotate the two solutions. The best matching factor after the rotations are given in Table 4.17. The two solutions are quite close. The comparison of the two approaches has been already discussed in Chapter 1 and 2.

Table 4.17: Sexual attitudes: Rotation of the factor loadings obtained from the programs LATENT & LISCOMP

	LATENT program		LISCOMP program	
Items	factor 1	factor 2	factor 1	factor 2
SEXLAW	0.420	-0.264	0.317	-0.905
GAYTEASC	0.901	-0.422	0.920	-0.356
GAYTEAH	0.907	-0.407	0.909	-0.403
GAYPUB	0.906	-0.319	0.837	-0.345
FGAYADPT	0.800	0.567	0.782	0.469
MGAYADPT	0.809	0.585	0.895	0.584
PMS	0.435	0.165	0.467	0.220
EXMS	0.300	0.120	0.326	0.150
SAME SEX	0.634	0.042	0.637	0.078

We decided not to give the scores for this data set due to the large number of response patterns involved.

4.3 Environment data, BSA 1991

The third data set is from the 1991 BSA survey. The questions analyzed here have been extracted from the environment section of the survey. The data set contains 7 binary and 7 continuous items. On this data set a two factor latent trait model has been fitted.

The third data set is from the environment section of the British Social Attitudes Survey in 1991. The mode of administration was self-completion. The question wording for the fourteen variables which have been extracted for the analysis are given in Appendix E. The sample contains 1079 individuals. The same items have been analyzed by Witherspoon and Martin (1992), in a paper with the characteristic title "What do we mean by green?". We use this data set to illustrate the fit of the mixed model on a bigger number of binary and continuous items and on a larger number of responses. We are interested in whether there is one underlying factor or more for the "green-ness" attitude. In other words if the "green-ness" attitude is more than one dimension then that means that people who are aware of some environmental issues may not be aware of other environmental issues. A way to test this assumption is to fit a latent variable model to these items, in which the latent variables will be indicators of the different dimensions of "green-ness" attitude. Witherspoon and Martin (1992) conducted a factor analysis and constructed three scales which have to do with people's willingness to take certain actions to protect or not to protect the environment. The first scale is the global green scale, the second is the pollution scale and the third is the nuclear power scale.

All fourteen items were based on 4 point scales and were treated as interval scale in earlier analysis. Most of these fourteen items are skewed and so by treating them all as continuous variables and then fit the linear factor model we violate the assumption about the normality of the observed variables. By dichotomizing some of them we certainly reduce the problem of misspecification of the model. What we do in this paper is to fit a two factor latent trait model on mixed items. We select 7 items to be treated as interval scale variables and 7 items to be treated as binary. To create the binary items we combine the first two categories (i.e 3 and 4) for each variable and recode them as "1" and we combine the last two categories (i.e 1 and 2) and recode them as "0".

Table 4.18 gives the maximum likelihood estimates of the two-factor latent trait model.

By looking the two- and three- way margins of the binary items after the mixed model has been fitted, we see that the two factor model has not much improved the fit on the margins compared to the one factor model. A considerable improvement in the fit from the two factor model has been made on the continuous part and that is observed by comparing the sample covariance matrix with the estimated covariance matrix for the one and two factor models.

The coefficients of the latent variables in Table 4.18 did not reveal the usual general factor. But if we disregard the negative signs of the second latent variable then we see that items 1, 2, 3, 4, 5, 6 and 7 load heavier on the first factor than the second factor and the opposite holds for the rest of the items, see also figure

Table 4.18: Environment section: Parameter estimates and standard errors for the two-factor latent trait model

Variable v_i	α_{i0}	α_{i1}	α_{i2}	π_i	α^*_{i1}	α^*_{i2}
insecticides [1]	2.72 (0.19)	1.46 (0.16)	-0.36 (0.16)	0.94	0.81	-0.20
ozone layer [2]	3.93 (0.41)	2.52 (0.34)	-0.49 (0.23)	0.98	0.92	-0.18
nuclear power [3]	1.92 (0.17)	1.46 (0.17)	-0.48 (0.15)	0.87	0.80	-0.27
lead from petrol [8]	3.10 (0.23)	0.82 (0.18)	-1.45 (0.18)	0.96	0.42	-0.75
industrial waste [9]	5.77 (0.67)	1.28 (0.33)	-1.43 (0.36)	1.00	0.59	-0.66
nuclear waste [10]	2.60 (0.21)	1.05 (0.20)	-1.24 (0.17)	0.93	0.55	-0.65
nuclear power [14]	0.96 (0.09)	0.76 (0.11)	-0.64 (0.11)	0.72	0.54	-0.45
Variable w_i	μ_i	λ_{i1}	λ_{i2}	Ψ_{ii}	λ^*_{i1}	λ^*_{i2}
greenhouse effect [4]	3.24 (0.04)	0.56 (0.03)	-0.17 (0.04)	0.35 (0.02)	0.67	-0.20
use earth's fuels [5]	3.21 (0.03)	0.49 (0.03)	-0.07 (0.04)	0.39 (0.02)	0.61	-0.08
species loss [6]	3.46 (0.03)	0.38 (0.03)	-0.09 (0.03)	0.35 (0.02)	0.53	-0.12
disposal of chem. [7]	3.60 (0.04)	0.35 (0.02)	-0.07 (0.03)	0.31 (0.02)	0.54	-0.10
industrial fumes [11]	3.46 (0.03)	0.18 (0.04)	-0.43 (0.02)	0.20 (0.01)	0.29	-0.67
acid rain [12]	3.36 (0.03)	0.23 (0.04)	-0.48 (0.03)	0.20 (0.01)	0.34	-0.69
aerosol damage [13]	3.31 (0.03)	0.26 (0.04)	-0.47 (0.03)	0.20 (0.01)	0.38	-0.67

4.6. The loadings for item 14 are moderate for both factors. Analytically, scale 1 contains the items: 1, 2, 3, 4, 5, 6, and 7 and scale 2 contains the items: 8, 9, 10, 11, 12, 13 and 14.

Our results are not directly comparable with Witherspoon et al. (1992) since we fit a two factor model and they fitted a three factor model. When we fit a two factor model using their method we find results very similar to ours. A general comment is that we find the green global scale (except from one item) plus two items of the nuclear power scale in our first scale and we find the pollution scale plus the rest of the items in the second scale. The items that have been extracted from the analysis suggest strongly that not only is there an effect which depends on the topic that each question asks but more than that there is a strong wording effect that has to do with the different things that questions 1-7 ask compared with questions 8-13. Questions 1-7 ask how concerned you are about certain environmental issues and questions 8-13 ask how serious you think an effect on our environment is from certain actions. That wording effect has as a result questions 1-7 to lie on one axis of the latent factor space and questions 8-13 to lie on the second axis. Our analysis shows that people respond differently to different environmental issues but this is also due to the fact that the question wording is different for the questions assigned to scale 1 and those assigned to scale 2.

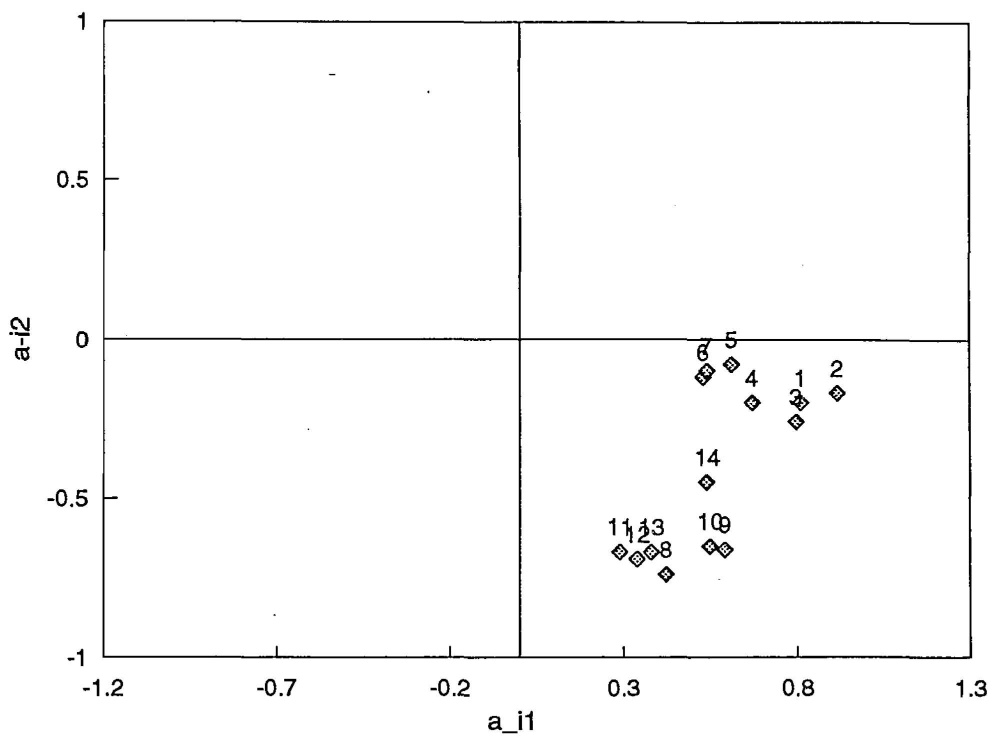

Figure 4.6: Environment data - Standardized factor loadings

We attempted to fit the same model with LISCOMP but it did not give us any result because of a program failure.

4.4 Analyses on simulated data

Data has been simulated according to the one and two factor latent trait model for mixed items, in order to provide sample estimates of the item parameters which can be compared with the known population values and so to evaluate the fit of the model to our empirical data. The method used is called parametric bootstrapping.

The simulated data are from the third data set presented here on sexual attitudes. For the one-factor model we generated a sample of N=1121 individuals from the standard normal distribution, $z \sim N(0,1)$. Given an individual's z-value, his responses to the r continuous items were generated according to the linear factor model, (chapter 1, section 1.2), and his responses to the s binary items according to the logistic item response model, (chapter 1, section 1.3.2). The parameters used are the ones estimated from the empirical data set, (BSA 1990). For the linear factor model, for each individual, an (e) value was sampled from the normal distribution with mean zero and variance Ψ_{ii}.

We analyzed the simulated data by fitting an one-factor latent trait model for mixed items. Table 4.19 gives the parameter estimates obtained under the correct model. Comparing the results with the ones we get from the empirical data, (Table 4.12), we see that they are very close. That indicates that the parameters were estimated quite accurately.

Table 4.19: Sexual attitudes: Simulated data - Parameter estimates for the one-factor latent trait model

Variable v_i	α_{i0}	α_{i1}	π_i	α^*_{i1}
SEXLAW [1]	1.65 (0.09)	0.60 (0.10)	0.84	0.52
GAYTEAS [2]	-0.77 (0.27)	6.83 (0.91)	0.32	0.99
GAYTEAH [3]	2.38 (3.75)	14.2 (16.5)	0.92	1.00
GAYPUB [4]	0.52 (0.15)	3.71 (0.32)	0.63	0.97
FGAYADP [5]	-2.62 (0.19)	2.01 (0.21)	0.07	0.89
MGAYADP [6]	-4.23 (0.41)	2.47 (0.34)	0.01	0.93
Variable w_i	μ_i	λ_{i1}	Ψ_{ii}	λ^*_{i1}
PMS [7]	3.56 (0.04)	0.54 (0.05)	1.90 (0.08)	0.36
EXMS [8]	1.62 (0.03)	0.17 (0.03)	0.68 (0.03)	0.20
SAME SEX [9]	2.01 (0.05)	0.90 (0.05)	1.62 (0.08)	0.58

Similarly with the one-factor model we can simulate data for the two-factor latent trait model for mixed items. The factor solution we get for the two-factor model is arbitrary since no constraints are imposed in the estimation of the mixed model. In order to investigate if there is a difference in the two factor solutions we get from the simulated and the empirical data we orthogonally rotated the two factor standardized solutions we got. The results are given in Table 4.20 and they are quite close.

Table 4.20: Sexual attitudes: Rotation of the standardized solution of the simulated and empirical data

	Simulated	data	Empirical	data
Variable v_i	α^*_{i1}	α^*_{i2}	α^*_{i1}	α^*_{i2}
SEXLAW [1]	0.37	0.19	0.43	0.26
GAYTEAS [2]	0.91	0.42	0.91	0.41
GAYTEAH [3]	0.89	0.43	0.92	0.39
GAYPUB [4]	0.93	0.27	0.91	0.30
FGAYADP [5]	0.79	-0.58	0.79	-0.58
MGAYADP [6]	0.81	-0.58	0.80	-0.60
Variable w_i	λ^*_{i1}	λ^*_{i2}	λ^*_{i1}	λ^*_{i2}
PMS [7]	0.43	-0.29	0.43	-0.17
EXMS [8]	0.28	-0.07	0.30	-0.13
SAME SEX [9]	0.59	-0.05	0.63	-0.06

Chapter 5

Missing values

Introduction

This chapter deals with the problem of missing values in attitude scales and the way these are treated in the analysis of mixed (binary and metric) manifest items with a latent trait model.

There are three types of missing data. The first type is noncoverage in which units are missing from the sampling frame, the second type is unit nonresponse in which all responses are missing for an individual in the sample and the third type is item nonresponse in which some of the responses are missing for an individual in the sample. The sources of unit nonresponse includes refusals or people not at home and the sources of item nonresponse includes refusals, don't know, interviewer error and response that was missed out. Here, we are interested in item non-response. All the different types of item nonresponse will be treated as one category coded as '9'.

The scope of our analysis is to include the missing values into the analysis of the manifest variables and to obtain information about the missing values based on what has been observed. Emphasis in our approach will be given not only to the estimation and the interpretation of the model coefficients but also to graphical methods based on posterior probabilities that will be used to obtain information about attitude from non expression of an opinion. The scoring of individuals on the latent factor space based on their response/ nonresponse pattern will be discussed.

Artificial data sets from Guttman, and mixed scales as well as a real data set from the BSA 1990 survey will be presented in order to examine the mechanism of the model and to illustrate its use. From the examples it will be apparent the use of the metric manifest variables in the analysis.

Multivariate data with missing values are analyzed either by disregarding the missing cases and carrying out the analysis on the complete data or by imputing the missing values. Imputation methods can be simple such as mean imputation, hot deck imputation, substitution and regression imputation. For a discussion of these methods see Little and Rubin (1987). Multiple imputation techniques have been also developed by Rubin (1987) in which more than one value for the missing items is imputed.

Another approach in the literature is that of maximum likelihood based methods in which a model is defined for the complete and the incomplete data. Little and Schluchter (1985) developed a ML estimation method for analysing mixed continuous and categorical data with missing values in the context of linear, logistic regression and discriminant analysis. The models discussed in their paper assume that the missing data are missing at random which means that the missingess depends only on observed variables. Rubin (1976) shows that if the data are missing at random then the likelihood based inference does not require the specification of a model for the missing data. In terms of Rubin's terminology this is an ignorable model with respect to the missing values.

Here, the work of Albanese and Knott (1992) for a latent variable model for binary items which allows for item non-response will be extended for the mixed model. Their method analyzes the response patterns as they are and estimates model parameters from a single analysis of the response / nonresponse patterns.

5.1 Models for non-response, binary variables

Albanese and Knott (1992) proposed three models for handling missing values in the analysis of attitudinal items with a latent trait model. These models do not distinguish between different sources of item non-response. They define a two dimension factor model in which one factor is named as attitude factor z_a and the other one is named as expression factor z_e. In other words z_a summarizes information about individual's attitude on a subject and z_e summarizes information about individual's propensity to express an opinion or not. Both factors, z_a, z_e are assumed to have independent standard normal distributions. Conditional on the two factors the responses (approve or disapprove) and the non-responses are taken to be independent, (conditional independence).

They also define probabilities of response and non-response to an item i together with probabilities of approval and disapproval of this item based on the breaking of the response function into two layers.

Before we go into defining these probabilities we should describe the basic idea of the Albanese and Knott (1992) approach.

Suppose that we have s binary items to analyze and there is a proportion of non response in each item. We create s pseudo items as follows, when an individual gives a response (approval or disapproval) then the pseudo item for this individual will take the value one, when an individual do not respond to this item then the pseudo item will take the value zero. At a next stage they fit a two factor model on the $(2 * s)$ items. So in a way the first s items provide us information about attitude and they are called attitudinal items and the next s items provide us information about expression and they are called expression items.

The response function is breaking into two layers.

For each binary item:

$$Pr(v_i = 1 \mid z_a, z_e, v_i \neq 9) = \pi_{ai}(z_a) \tag{5.1}$$

$$Pr(v_i \neq 9 \mid z_a, z_e) = \pi_{ei}(z_a, z_e) \tag{5.2}$$

It follows that,

$$Pr(v_i = 1 \mid z_a, z_e) = \pi_{ai}(z_a)\pi_{ei}(z_a, z_e)$$

$$Pr(v_i = 0 \mid z_a, z_e) = (1 - \pi_{ai}(z_a))\pi_{ei}(z_a, z_e)$$

$$Pr(v_i = 9 \mid z_a, z_e) = 1 - \pi_{ei}(z_a, z_e)$$

The above formulation of the response function indicates that if an individual responds, (equation 5.1), then the expressed attitude is not dependent on z_e but the probability that an individual does respond, (equation 5.2), depends on both z_a and z_e, where z_e is individuals inherent responsiveness for all questions. In other words individuals with high z_a may have a different probability of not responding than individuals with low z_a.

A brief description of the three models proposed by Albanese and Knott (1992) will be given here.

Model 1 is a simple model, in which the probability of an individual to express or not express an opinion is constant across individuals and independent of other items. It is written as:

$$\text{logit}(\pi_{ai}(z_a)) = \alpha_{i0} + \alpha_{i1} z_a$$

$$\text{logit}(\pi_{ei}(z_e)) = e_{i0}$$

where $i = 1, \cdots, s$

Model 2 allows the probability of expressing or not expressing an opinion from an individual to depend on an expression factor. Probability of responding varies by individual but is independent of attitude. This model does not give us any information of how attitude influences expression. It is written as:

$$\text{logit}(\pi_{ai}(z_a)) = \alpha_{i0} + \alpha_{i1} z_a$$

$$\text{logit}(\pi_{ei}(z_e)) = e_{i0} + e_{i1} z_e$$

where $i = 1, \cdots, s$

Model 3 is a more sophisticated one in which the probability of responding differs by individual and may depend on the individual's attitude. This model has been looked at thoroughly as it will be shown in the following sections.

The model is written as:

$$\text{logit}(\pi_{ai}(z_a)) = \alpha_{i0} + \alpha_{i1} z_a, \qquad i = 1, \cdots, s$$

$$\text{logit}(\pi_{ei}(z_a, z_e)) = e_{i0} + e_{i1} z_a + e_{i2} z_e, \qquad i = 1, \cdots, s$$

the coefficient e_{i1} shows how the log of the odds of expression of an opinion increases or decreases with respect to the attitude dimension. Models 1, 2 and 3 have been fitted on seven binary items on abortion, (see Knott, Albanese, and Galbraith 1990), using the program TWOMISS. In this paper the parameter estimates and the scoring of the individuals have been reported. They reported that model 3 gave the best fit compared to model 1 and 2. Model 3 is the one that is going to be looked at for the mixed case.

5.2 Model for non-response, mixed variables

The idea described above for binary items will be extended here for mixed manifest variables.

Similarly for the case of mixed items the pseudo items will be take the value zero if the individual did not respond to this item and 1 otherwise. After we have created the pseudo items there will be a number of $(2*s+r)$ binary items and r continuous items. We proceed by fitting a two factors latent trait model on the $2*(r+s)$ items.

Equivalent to the results presenting in the previous section we can break the response function into two layers.

For each attitude binary item:

$$Pr(v_i = 1 \mid z_a, z_e, v_i \neq 9) = \pi_{ai}(z_a), \quad i = 1, \cdots, s \tag{5.3}$$

For each attitude continuous item:

$$(w_i \mid z_a, z_e, w_i \neq 9) \sim N(\mu_i + \lambda_{i1} z_a, \Psi_{ii}) \quad i = 1, \cdots, r \tag{5.4}$$

For each expression item:

$$Pr(x_i \neq 9 \mid z_a, z_e) = \pi_{ei}(z_a, z_e) \quad i = 1, \cdots, r+s \tag{5.5}$$

Where for our model:

$$\text{logit}\pi_{ai}(z_a) = \alpha_{i0} + \alpha_{i1} z_a$$

and

$$\text{logit}\pi_{ei}(z_a, z_e) = e_{i0} + e_{i1} z_a + e_{i2} z_e$$

It follows that

$$Pr(v_i = 1 \mid z_a, z_e) = \pi_{ai}(z_a)\pi_{ei}(z_a, z_e)$$

$$Pr(v_i = 0 \mid z_a, z_e) = (1 - \pi_{ai}(z_a))\pi_{ei}(z_a, z_e)$$

$$Pr(v_i = 9 \mid z_a, z_e) = 1 - \pi_{ei}(z_a, z_e)$$

$$f(w_i \mid z_a, z_e) = N(\mu_i + \lambda_{i1} z_a, \Psi_{ii}) * \pi_{ei}(z_a, z_e)$$

$$Pr(w_i = 9 \mid z_a, z_e) = 1 - \pi_{ei}(z_a, z_e)$$

5.2.1 Estimation of the model

We make two assumptions for this model. The first one is the known assumption of conditional independence which says that the responses (approve or disapprove) or non-responses to the $(2 * (r + s))$ items are independent given the vector of latent variables (z_a, z_e). The second assumption says that given that an individual responded to an item the probability of approve or disapprove does not depend on the factor z_e, (see equations 5.3 and 5.4).

The density function to be looked at is the joint distribution of the manifest variables which under the assumption of conditional independence is given by:

$$f(\mathbf{v}_h, \mathbf{w}_h) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\mathbf{v}_h \mid \mathbf{z}) g(\mathbf{w}_h \mid \mathbf{z}) h(\mathbf{z}) d\mathbf{z} \tag{5.6}$$

where $\mathbf{v}_h, \mathbf{w}_h$ represents the responses to the $2 * (r + s)$ manifest variables of the h^{th} individual and $h(\mathbf{z})$ denotes the prior distribution of the latent variables, z_a and z_e taken to be independent standard normal.

The conditional distribution of $\mathbf{w}_h \mid \mathbf{z}$ is given by:

$$g(\mathbf{w}_h \mid \mathbf{z}) = \prod_{i=1}^{r} [N(\mu_i + \lambda_{i1} z_a, \Psi_{ii}) \pi_{e,i+p^*}(\mathbf{z})]^{w_{i+p^*,h}} [(1 - \pi_{e,i+p^*}(\mathbf{z})]^{1-w_{i+p^*,h}} \tag{5.7}$$

where, $p^* = r + 2s$

The conditional distribution of $\mathbf{v}_h \mid \mathbf{z}$ is given by:

$$g(\mathbf{v}_h \mid \mathbf{z}) = \prod_{i=1}^{s} [\pi_{ai}(z_a)^{v_{ih}} (1 - \pi_{ai}(z_a))^{1-v_{ih}} \pi_{e,i+p}(\mathbf{z})]^{v_{i+p,h}} [1 - \pi_{e,i+p}(\mathbf{z})]^{1-v_{i+p,h}} \tag{5.8}$$

where, $p = r + s$

The log-likelihood for a random sample of size n is

$$L = \sum_{h=1}^{n} \log f(\mathbf{v}_h, \mathbf{w}_h) = \sum_{h=1}^{n} \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\mathbf{w}_h \mid \mathbf{z}) g(\mathbf{v}_h \mid \mathbf{z}) h(\mathbf{z}) d\mathbf{z} \tag{5.9}$$

The log-likelihood is maximized using the EM algorithm described in Chapter 2. The model can be fitted with the program LATENT. The steps of the fit are given here. First we generate the $r + s$ pseudo items from the original manifest variables, secondly we fit the two factor model on the $2 * (r + s)$ items treating the $2 * s + r$ variables , (r here refers to the pseudo binary items derived from the metric variables), as binary and the original metric variables r as metric constrain the loadings of the second factor for the manifest variables, binary and metric, to zero.

5.2.2 Interpretation of the model

From the formulation of the model presented in the section above we see that the missing values are included in the analysis with the observed values and parameter estimates are obtained from a single analysis of both the missing and the observed values.

This model allows the probability of expressing an opinion to depend on two factors (z_a, z_e) and the probability of approving or disapproving of an item given that an individual responded to that item to depend only on one factor z_a.

By fitting this model on the $2 \times (r + s)$ mixed items we are interested in investigating how attitude affects expression. This information can be obtained from the coefficient e_{i1} which measures the effect of the attitude on the log odds of the response probability.

But it is not enough to look only the magnitude and the sign of these coefficients. What we are mostly interested in is to find a way to obtain information about attitude from non-expression. That is explored in the following section.

5.2.3 Posterior analysis

From the model parameters we obtain information on how attitude affects expression and also information on how likely or unlikely is to get a response for an item. But we are also interested in obtaining information about the missing values and what they represent in our sample. We propose here to look for each item at the posterior distribution of the attitude latent variable z_a given the possible responses for that item. So for binary items we are interested in observing the relative position of the $h(z_a \mid v_i = 9)$ with respect to $h(z_a \mid v_i = 0)$ and $h(z_a \mid v_i = 1)$ and for metric variables we look at the relative position of $h(z_a \mid w_i = 9)$ with respect to the three quartiles.

These posterior probabilities can be computed after we estimated the model as follows:

$$h(z_a \mid v_i = k) = \frac{\int_{-\infty}^{\infty} g(v_i = k \mid z_a, z_e) h(z_a) h(z_e) dz_e}{f(v_i = k)} \tag{5.10}$$

$k = 0, 1, 9$

where the form of the $g(v_i = k \mid z_a, z_e)$ is given in equation (5.8) and

$$f(v_i = k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(v_i = k \mid z_a z_e) h(z_a) h(z_e) dz_a dz_e$$

and for the metric variables:

$$h(z_a \mid W_i = w_i) = \frac{\int_{-\infty}^{\infty} g(W_i = w_i \mid z_a, z_e) h(z_a) h(z_e) dz_e}{f(W_i = w_i)} \tag{5.11}$$

where the form of the $g(W_i = w_i \mid z_a, z_e)$ is given in equation (5.7).

For the case $w_i = 9$

$$Pr(W_i = 9 \mid z_a, z_e) = 1 - \pi_{ei}(z_a, z_e)$$

5.3 Applications

In this section we will present the results we found when we fitted model 3 on a number of data sets with missing values. All except for one of the data sets used here are artificially created for illustrating reasons. The reason we are looking first at some artificial examples in which a specific number of patterns occur is because it is easier to obtain information about the mechanism of the model. An interesting result emerging from the analysis of the data is that the metric variables provide information that some times reduces the indeterminacy that arises from some response patterns and increases the strength of the predictive scope of the model with respect to the missing values. The posterior analysis presented above will be looked at for all the data sets. We are first interested in using the model to score the missing value for an item relative to the other responses for that item and to use this information to rank individuals on the attitude latent dimension.

5.3.1 Guttman and non-type scale variables

First our plan is to fit the model presented above to a number of items that form a perfect Guttman scale. The reason is that in a Guttman scale we can predict in a satisfactory degree the outcome of our analysis and so we are able up to a degree to validate our model. Although Guttman scales have a deterministic nature as far as concern the structure of the responses from the individuals it is worth examined here for understanding the mechanism of the model. It is quite probable that the model does not fit well Guttman scale items for the reason that the response function of each item behaves as a threshold, that is actually verified by the large parameter estimates for the discrimination coefficients of the attitude items. The discrepancies in the observed and expected two- and three-way margins of the attitudinal and the expression (pseudo) items will be looked at for the data analyzed here. These discrepancies are measured with the statistic given by $(O-E)^2/E$.

However, the Guttman scale is used here only for illustration purposes and the results should be looked at with cautious.

The data set used consists of 4 binary items and one metric item. The four binary items construct a perfect Guttman scale and the metric item is highly correlated with the scale. Three different experiments will be looked at here. The first one consists of only the four binary items and there is a number of missing values on the fourth item of the response pattern 1 1 1 1. So the effect of a single item with missing values will be examined. The response patterns together with the frequency for each pattern are given in Table (5.1).

Table 5.1: Guttman scale 1

Response pattern	frequency
1 1 1 1	70
1 1 1 9	70
1 1 1 0	70
1 1 0 0	70
1 0 0 0	70
0 0 0 0	70

When we fit the model on this data we expect that the posterior distribution

$h(z_a \mid v_4 = 9)$ will be somewhere in the middle of $h(z_a \mid v_4 = 0)$ and $h(z_a \mid v_4 = 1)$. The reason is that we expect that the model extracts information from the response patterns in order to place the missing value. This indeterminacy is created because the '9' could come either from the response pattern 1 1 1 1 or 1 1 1 0. Figure (5.1) shows that the model places the $h(z_a \mid v_4 = 9)$ closer or even above 1. That is actually surprisingly since we would expect 9 to be between 0 and 1. The fit of the model on the margins looks satisfactory. One reason for that result might be that '9's come only in one response pattern and so the absence of clustering of '9' within other response patterns does not help the model to place '9' as it is expected.

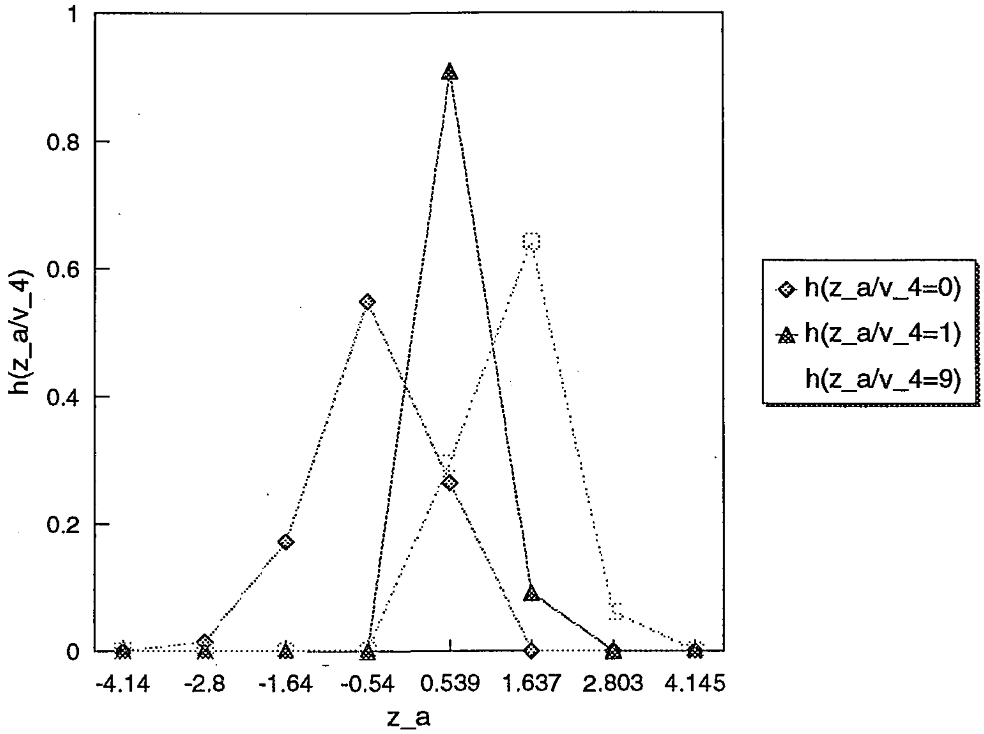

Figure 5.1: Guttman scale 1, posterior probabilities

The second experiment consists of four binary items and there is a number of missing values in the fourth item of the response pattern 1 1 0 0. The response patterns are given in Table (5.2).

From the analysis of this response patterns we would expect no indeterminacy due to the fact that the response pattern 1 1 0 9 can only come from 1 1 0 0. The discrepancy measure show a good fit on the expression items and a less good fit on the attitude items. Looking at figure (5.2) we see that there is no indeterminacy here and the $h(z_a \mid v_4 = 9)$ is correctly placed closer to 0.

Table 5.2: Guttman scale 2

Response pattern	frequency
1 1 1 1	70
1 1 1 0	70
1 1 0 0	70
1 1 0 9	70
1 0 0 0	70
0 0 0 0	70

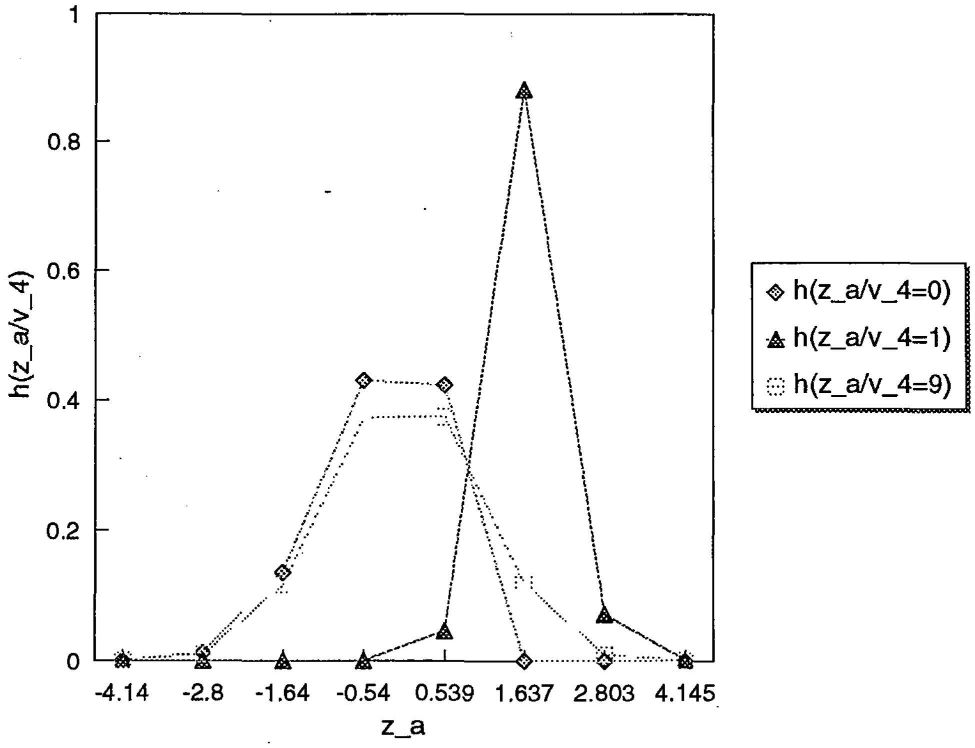

Figure 5.2: Guttman scale 2, posterior probabilities

The third experiment consists of four binary variables and one metric variable. The metric variable was chosen to be highly correlated with the binary items. The reason we add the metric variable here is to see whether the metric variable can reduce the indeterminacy created from the response pattern 1 1 1 1 and 1 1 1 0 with respect to the missing value '9. The response patterns are given in Table (5.3).

Table 5.3: Guttman mixed scale

Response pattern	frequency
1 1 1 1 5	70
1 1 1 9 4	50
1 1 1 0 4	70
1 1 0 0 3	70
1 0 0 0 2	70
0 0 0 0 1	70

By looking at figure (5.3) we see that the indeterminacy has disappeared and that $h(z_a \mid v_4 = 9)$ has correctly been placed closer to 0 since the response pattern 1 1 1 9 4 can only now come from 1 1 1 0 4. Small discrepancies are observed in the margins of the expression items which shows a good fit.

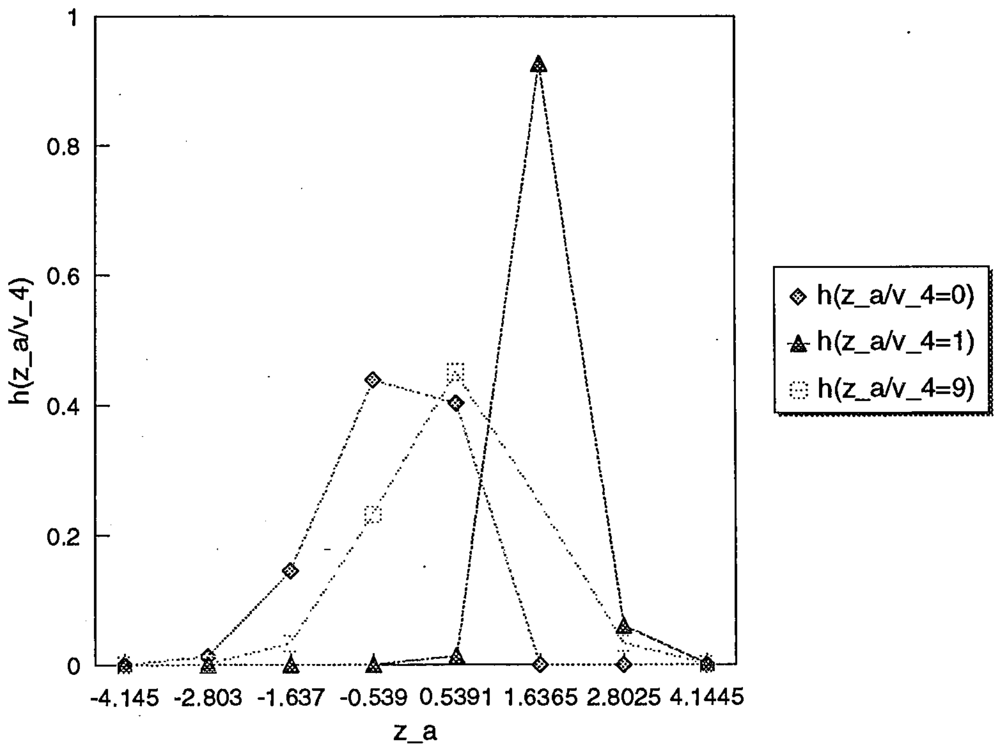

Figure 5.3: Guttman mixed scale, posterior probabilities

We are also going to investigate the case where the scale is not a perfect Guttman scale but a non-scale type. The data set used is given in Table (5.4).

Table 5.4: Non-scale type 1

Response pattern	frequency
1 1 1 1	70
1 1 1 0	70
1 1 0 1	70
1 1 0 0	70
1 1 0 9	70
1 0 0 0	70
0 0 0 0	70

We added the response pattern 1 1 0 1 and missing values occur on the fourth item of the response pattern 1 1 0 1. That will create an indeterminacy again because the response pattern 1 1 0 9 can come either from pattern 1 1 0 0 or 1 1 0 1. Looking at figure (5.4) we see that the $h(z_a \mid v_4 = 9)$ is somewhere in the middle of 0 and 1. In that example the model reflects the indeterminacy problem and places the missing value in the middle of 0 and 1. Comparing the margins on the expression items for that example with the margins for the first Guttman scale example these ones look marginally better.

However this indeterminacy dissapears, (see figure 5.5), when a strong correlated with the binary items metric variable is included. The data set is given in Table (5.5).

Table 5.5: Non-scale type 2

Response pattern	frequency
1 1 1 1 5	70
1 1 1 0 4	70
1 1 0 1 4	70
1 1 0 0 3	70
1 1 0 9 3	70
1 0 0 0 2	70
0 0 0 0 1	70

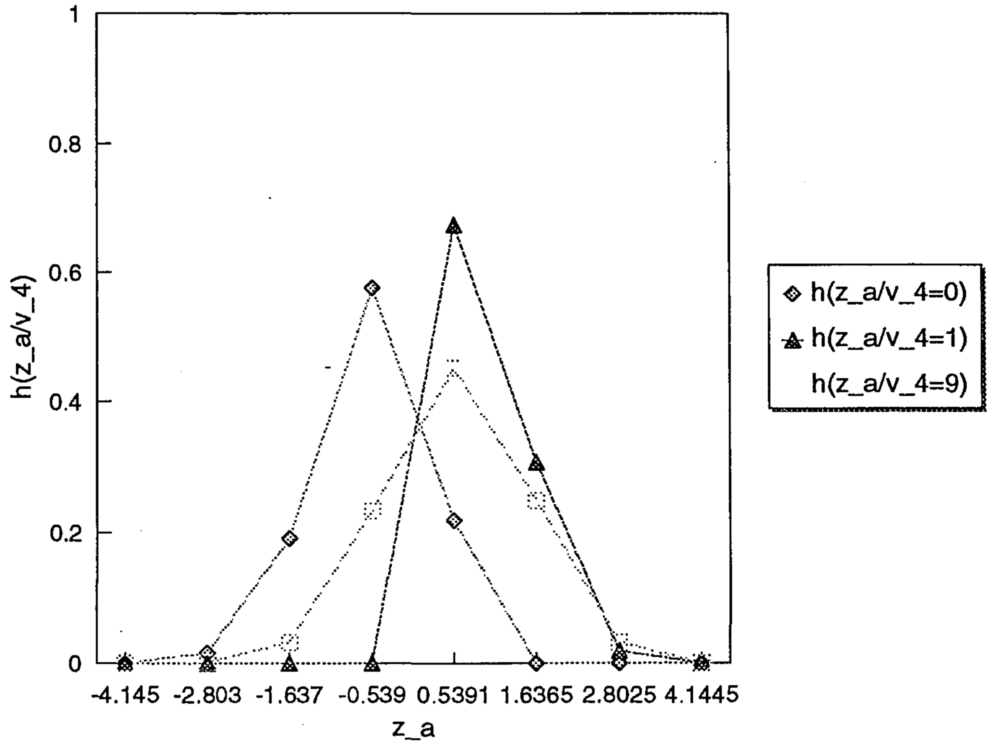

Figure 5.4: Non-scale type 1, posterior probabilities

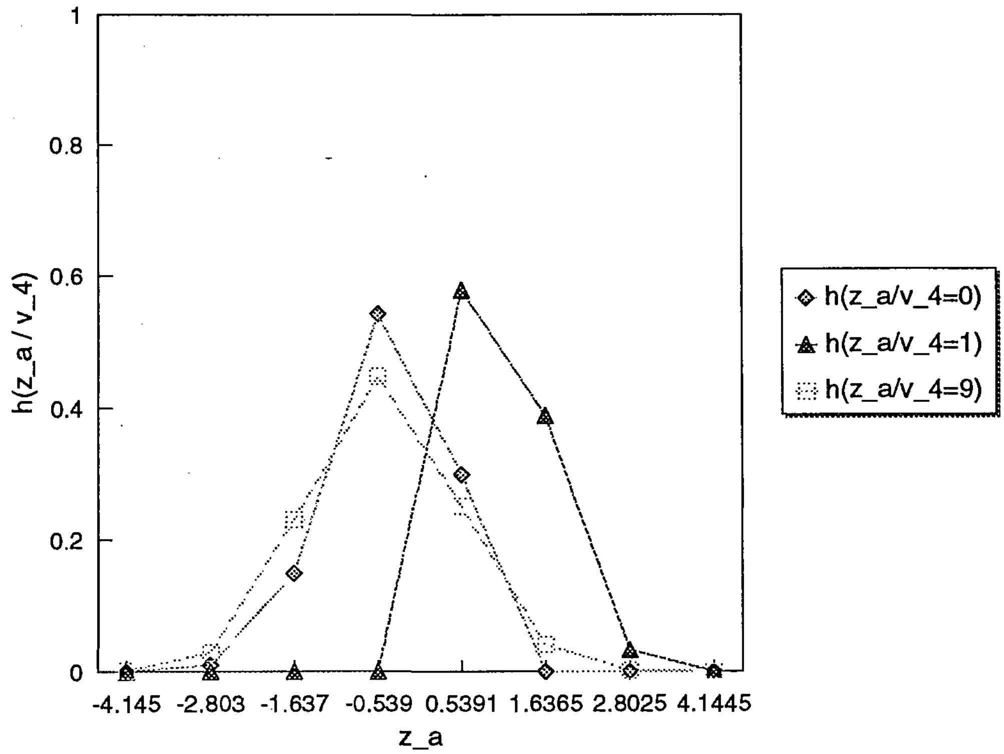

Figure 5.5: Non-scale type 2, posterior probabilities

5.3.2 Mixed scale variables

In this section we look at response patterns with binary and metric variables. The metric variables can be seen as Likert scales in which the respondent has to choose among several response categories, indicating various strengths of agreement and disagreement. Likert scales have been widely used in factor analysis in which interest is centered in the examination of the underlying structure of the set of manifest items.

Here, Likert scale items together with binary items will be investigated in the case of item non-response. The same analysis as with the Guttman scale will be used.

The first example will look at is one with four binary items and one likert item which here will be treated as metric. The response patterns together with their frequencies are given in Table (5.6). The correlation matrix for this data set has been computed with the program PRELIS, (Jöreskog and Sörbom 1988), and are given in Table (5.7).

Table 5.6: Mixed scale 1

Responses	freq	Responses	freq
0 0 0 1	50	1 1 0 4	40
0 0 0 2	50	0 1 1 2	40
0 0 1 2	8	0 1 1 4	40
0 0 1 3	4	1 1 1 2	30
0 1 0 3	25	1 1 1 3	40
1 0 1 2	40	1 1 1 4	50
1 1 0 2	10	1 1 1 5	60
1 1 0 3	40	1 1 1 9	25

Table 5.7: Correlation matrix 1

	bin1	bin2	bin3	cont1
bin1	1.00			
bin2	0.66	1.00		
bin3	0.47	0.53	1.00	
cont1	0.60	0.95	0.45	1.00

We would probably expect that the posterior distribution of z_a given $W = 9$ will be placed somewhere between 3 and 4. By looking at figure (5.6) we see that 9 is placed above 5. The model cannot predict the place of 9 correctly probably due to the fact that 9's come only with the same response pattern (1 1 1 9) and there is no clustering of 9's within other patterns.

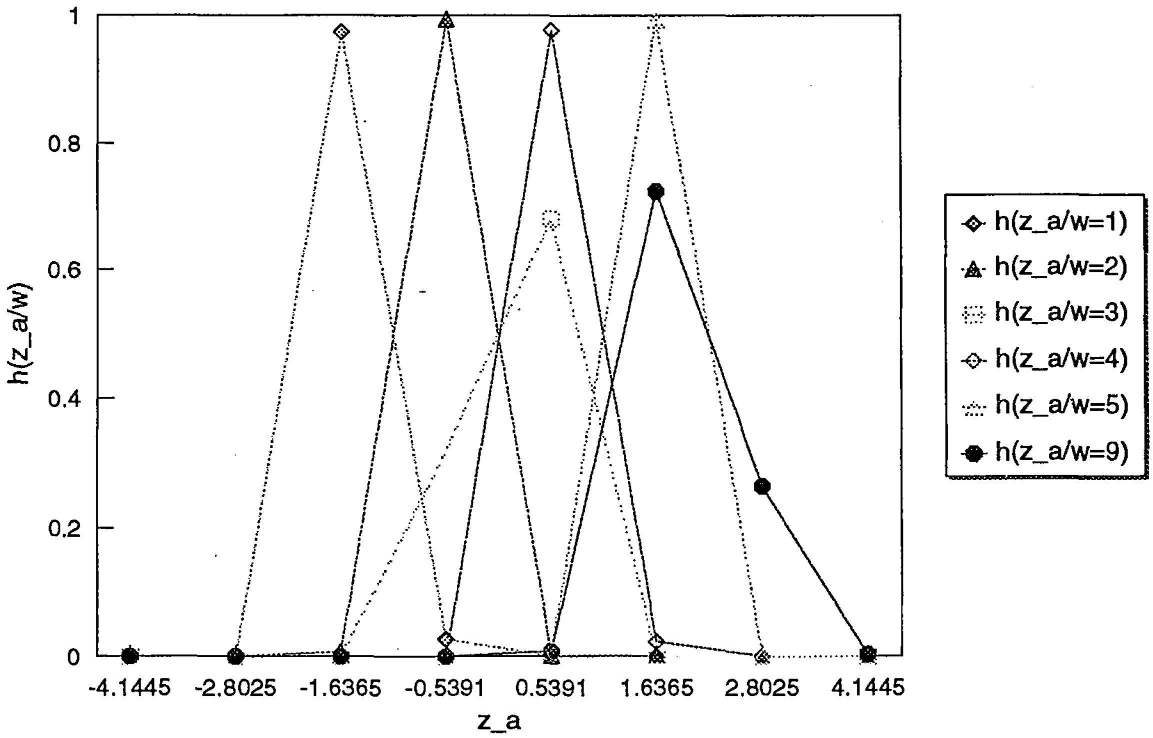

Figure 5.6: Mixed scale 1, posterior probabilities

It is also interesting to look at the posterior mean of the attitude latent variable given the response pattern of each individual, $E(z_a \mid \mathbf{w}_h \mathbf{v}_h)$. These results are given in Table (5.8). The pattern 1 1 1 9 scores higher than any other pattern.

For the second example we used the same response patterns as in Table (5.6) by instead of taking 25 cases of 1 1 1 9 we take 25 cases of 0 0 0 9. The correlation matrix for this data set is given in Table (5.9). The results are given in figure (5.7). The response 9 for the fourth item is placed below 0. So again here the model did not work as we expected. The posterior mean for the individuals is given in Table (5.10)

As a third example we used the same response patterns as in example 1 and 2 but we included 25 cases of 1 1 1 9 and 25 cases of 0 0 0 9. The correlation matrix for this data set is given in Table (5.11). In that case the model correctly placed 9 somewhere between 2 and 4 as it can be seen in figure (5.8). The ranking of the

Table 5.8: Posterior mean: mixed scale 1

$E(z_a \mid \mathbf{w}, \mathbf{v})$	responses
-1.63	0 0 0 1
-0.58	0 0 0 2
-0.54	1 0 1 2
-0.54	0 0 1 2
-0.54	0 0 1 3
-0.54	0 1 1 2
-0.54	1 1 0 2
-0.54	1 1 1 2
-0.16	0 1 0 3
0.31	1 1 0 3
0.46	1 1 1 3
0.54	0 1 1 4
0.55	1 1 0 4
0.58	1 1 1 4
1.63	1 1 1 5
1.97	**1 1 1 9**

Table 5.9: Correlation matrix 2

	bin1	bin2	bin3	cont1
bin1	1.00			
bin2	0.69	1.00		
bin3	0.50	0.57	1.00	
cont1	0.60	0.95	0.45	1.00

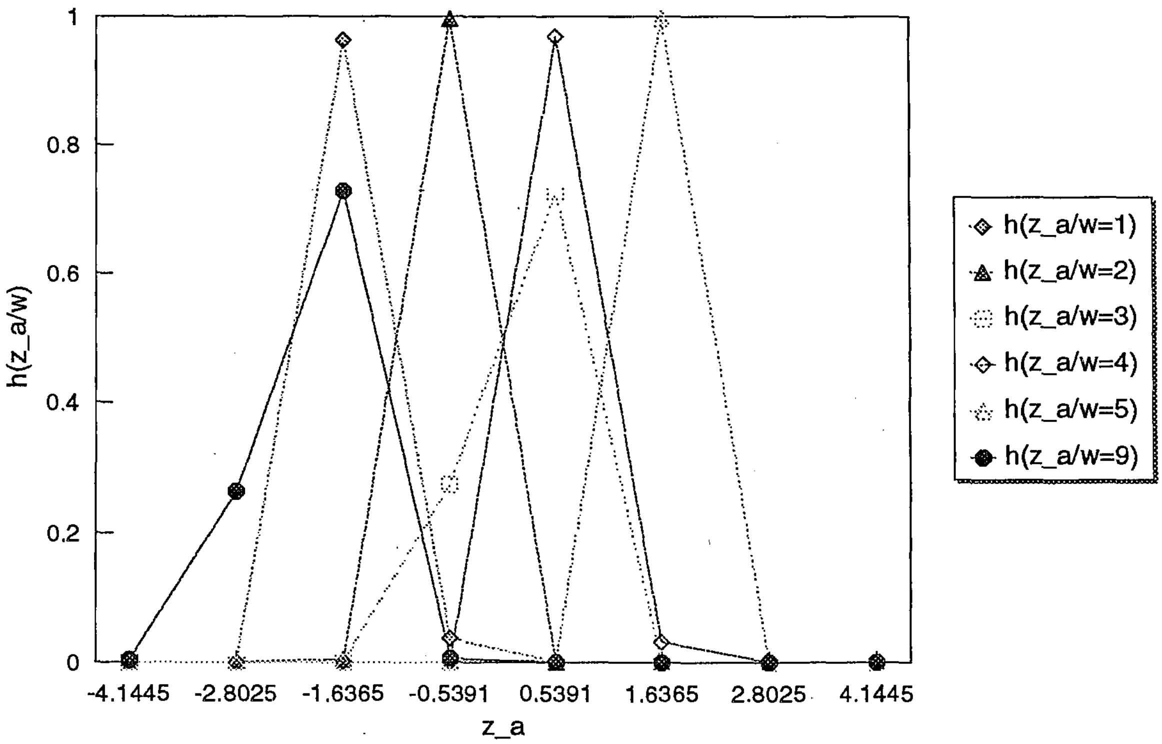

Figure 5.7: Mixed scale 2, posterior probabilities

Table 5.10: Posterior mean: mixed scale 2

$E(z_a \mid \mathbf{w}, \mathbf{v})$	responses
-2.01	**0 0 0 9**
-1.63	0 0 0 1
-0.56	0 0 0 2
-0.54	0 0 1 2
-0.54	1 0 1 2
-0.54	0 0 1 3
-0.54	0 1 1 2
-0.54	1 1 0 2
-0.54	1 1 1 2
-0.05	0 1 0 3
0.38	1 1 0 3
0.48	1 1 1 3
0.55	0 1 1 4
0.56	1 1 0 4
0.59	1 1 1 4
1.63	1 1 1 5

individuals is given in Table (5.12). The ranking is satisfactory because the response pattern 0 0 0 9 is between 0 0 0 1 and 0 0 0 2 and the response pattern 1 1 1 9 is somewhere between 1 1 1 3 and 1 1 1 5.

Table 5.11: Correlation matrix 3

	bin1	bin2	bin3	cont1
bin1	1.00			
bin2	0.71	1.00		
bin3	0.53	0.59	1.00	
cont1	0.60	0.95	0.45	1.00

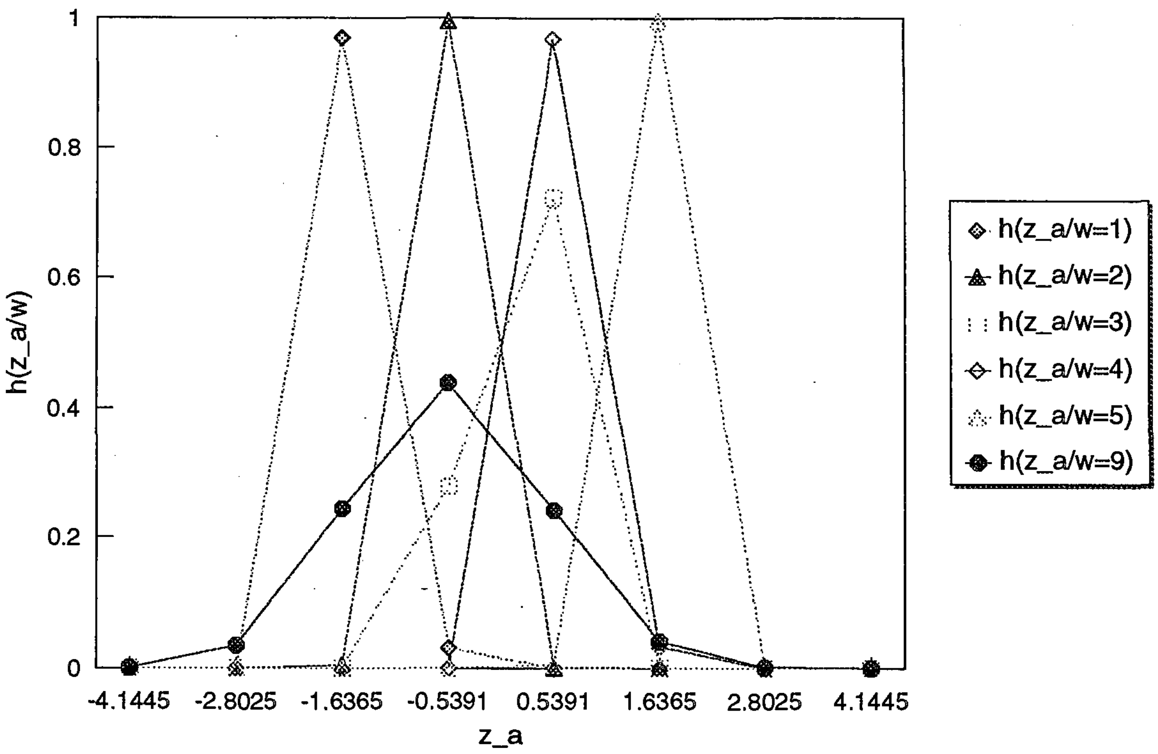

Figure 5.8: Mixed scale 3, posterior probabilities

We are now going to look at the case where we have two Likert scale variables rather than one. The response patterns are given in Table (5.14) and the correlation matrix for this data set is given in Table (5.13).

Looking at figure (5.9) we see that the second metric variable reinforced the model to predict better the place of 9 here between 3 and 4. From table (5.15) we see that response pattern 1 1 1 9 2 scores higher than 1 1 1 2 2 but lower than 1 1 1 3 3 and the response pattern 1 1 1 9 3 scores lower than 1 1 1 3 3 but above 1 1 1 3 2.

Table 5.12: Posterior mean: mixed scale 3

$E(z_a \mid \mathbf{w}, \mathbf{v})$	responses
-1.63	0 0 0 1
-1.51	**0 0 0 9**
-0.56	0 0 0 2
-0.55	0 0 1 2
-0.54	1 0 1 2
-0.54	0 0 1 3
-0.54	0 1 1 2
-0.54	1 1 0 2
-0.54	1 1 1 2
-0.11	0 1 0 3
0.36	1 1 0 3
0.48	1 1 1 3
0.53	**1 1 1 9**
0.55	0 1 1 4
0.55	1 1 0 4
0.59	1 1 1 4
1.63	1 1 1 5

Table 5.13: Correlation matrix 4

	bin1	bin2	bin3	cont1	cont2
bin1	1.00				
bin2	0.66	1.00			
bin3	0.47	0.53	1.00		
cont1	0.60	0.95	0.45	1.00	
cont2	0.41	0.86	0.23	0.88	1.00

Table 5.14: Mixed scale 4

Responses	freq	Responses	freq
0 0 0 1 1	50	1 0 1 2 2	40
0 0 0 2 2	50	0 1 1 2 2	40
0 0 1 2 2	3	0 1 1 4 4	40
0 0 1 2 3	5	**1 1 1 2 2**	30
0 0 1 3 2	1	**1 1 1 3 2**	32
0 0 1 3 3	3	**1 1 1 3 3**	8
0 1 0 3 3	9	**1 1 1 4 3**	50
0 1 0 3 4	16	1 1 1 5 4	10
1 1 0 2 3	10	1 1 1 5 5	50
1 1 0 3 3	40	**1 1 1 9 2**	10
1 1 0 4 4	40	**1 1 1 9 3**	15

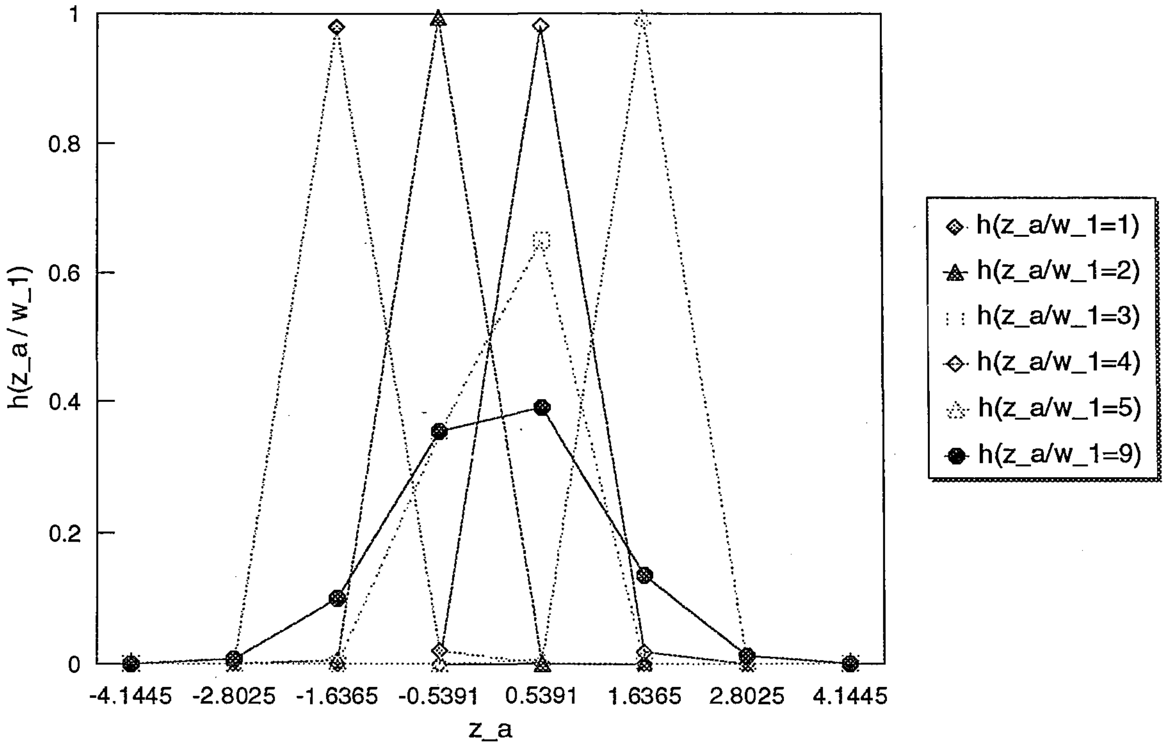

Figure 5.9: Mixed scale 4, posterior probabilities

Table 5.15: Posterior mean: mixed scale 4

$E(z_a \mid \mathbf{w}, \mathbf{v})$	responses
-1.64	0 0 0 1 1
-0.54	0 0 0 1 2
-0.54	1 0 1 2 2
-0.54	0 0 1 2 3
-0.54	0 1 1 2 2
-0.54	1 1 1 2 2
-0.54	0 0 1 3 3
-0.54	1 1 0 2 3
-0.53	**1 1 1 9 2**
-0.53	1 1 1 3 2
0.35	0 1 0 3 3
0.49	1 1 0 3 3
0.50	**1 1 1 9 3**
0.51	1 1 1 3 3
0.54	0 1 0 3 4
0.54	1 1 1 4 3
0.54	0 1 1 4 4
0.54	1 1 0 4 4
1.63	1 1 1 5 4
1.64	1 1 1 5 5

5.3.3 A real example: BSA 1990

The data set used here has been extracted from the British Social Attitudes, 1990, Survey. There were 1270 individuals who were asked questions on sexual relationships. The questions given below are a subset of the questions analyzed already in Chapter 4.

1...Now I would like you to tell me whether, in your opinion, it is acceptable for a homosexual person to be a teacher at a school? [GAYTEASC]

2...Now I would like you to tell me whether, in your opinion, it is acceptable for a homosexual person to be a teacher in a college or a university? [GAYTEAHE]

3...Now I would like you to tell me whether, in your opinion, it is acceptable for a homosexual person to hold a responsible position in public life? [GAYPUB]

4...What about sexual relations between two adults of the same sex? [SAME SEX]

If we exclude the responses "depends/varies", "don't know" and "not answered" from the above items the sample size reduces to 1215 individuals. The items 1 to 3 are binary items with response categories 1 for agree and 0 for disagree and item 4 is a five point scale item with responses "always wrong", "mostly wrong", "sometimes wrong", "rarely wrong" and "not wrong at all", treated here as discrete. The percentage of non response for item 1 is 1.8% for item 2 is 1.9% for item 3 is 2% and for item 4 is 0.8%.

First we fit a one factor model to the four items, excluding the missing values. The results are given in Table (5.16).

Table 5.16: Parameter estimates and standard errors for the one-factor latent trait model, complete data

Variable v_i	α_{i0}	α_{i1}	π_i	α^*_{i1}
GAYTEASC [1]	-0.82 (0.42)	9.73 (1.68)	0.31	0.99
GAYTEAHE [2]	1.55 (0.76)	11.15 (3.19)	0.82	0.99
GAYPUB [3]	0.80 (0.16)	3.53 (0.29)	0.69	0.97
Variable w_i	μ_i	λ_{i1}	Ψ_i	λ^*_{i1}
SAME SEX [4]	2.03 (0.06)	0.84 (0.06)	0.57 (0.07)	0.56

From the tables of the one- two- and three-way margins we see that the one factor model fits the data well. The π_i column shows that item 1 has a very low probability of receiving a positive response from the median individual. The item SAME SEX as well has a relative low mean score (2.03).

Table (5.17) gives the parameter estimates of the mixed model for missing values.

Table 5.17: Parameter estimates and standard errors for the two-factor latent trait model with missing values

Variable v_i	α_{i0}	α_{i1}	α_{i2}	π_i	α^*_{i1}	α^*_{i2}
GAYTEAS [1]	-1.36 (0.35)	6.03 (0.64)	0.00 (1.00)	0.20	0.99	0.00
GAYTEAH [2]	0.98 (0.28)	6.01 (0.52)	0.00 (1.00)	0.73	0.99	0.00
GAYPUB [3]	0.61 (0.11)	3.05 (0.19)	0.00 (1.00)	0.65	0.95	0.00
Variable w_i	μ_i	λ_{i1}	λ_{i2}	Ψ_i	λ^*_{i1}	λ^*_{i2}
SAME SEX [4]	1.98 (0:05)	0.83 (0.05)	0.00 (1.00)	1.56 (0.04)	0.55	0.00
Variable v_i	e_{i0}	e_{i1}	e_{i2}	π_i	e^*_{i1}	e^*_{i2}
GAYTEAS [1]	24.73 (****)	-0.08 (0.61)	-13.3 (****)	1.00	-0.00	-0.99
GAYTEAH [2]	11.4 (4.41)	0.94 (0.49)	-5.16 (2.59)	1.00	0.18	-0.97
GAYPUB [3]	6.40 (0.94)	1.20 (0.49)	-2.22 (0.51)	0.99	0.44	-0.82
SAME SEX[4]	6.71 (1.37)	-1.04 (0.56)	-1.69 (0.69)	0.99	-0.46	-0.76

From the tables with the one- two- and three-way margins the fit of the model looks satisfactory. As we have already said the formulation of the model allows attitude to affect expression. This information can be obtained by looking at the coefficients e_{i1}. The values of this coefficients will be discussed in connection with the posterior probabilities for the four items given below. Item 1 has a value of $e_{i1} = -0.08$ that indicates that attitude is not related to expression and as a result no information can be obtained for attitude from non-expression. From the posterior analysis discussed above we find that the $h(z_a \mid v_i = 9)$ for item 1 is in the middle of 0 and 1 (see figure 5.10). Item 2 has a value of $e_{i1} = 0.94$ that indicates that the more positive attitude an individual has towards homosexuality the more chances he has to respond and since he did not respond it is more likely that he will be on the left side of the attitude scale, (see figure 5.11). Item 3 is even more closer to 0 since the value of $e_{i1} = 1.20$, (see figure 5.12). Lastly for item 4 the value of $e_{i1} = -1.04$ and that indicates that the more positive attitide an individual has towards to homosexuality the less chances he has to respond and since he did not respond it is more likely to be on the wright part of the attitude scale, see figure 5.13.

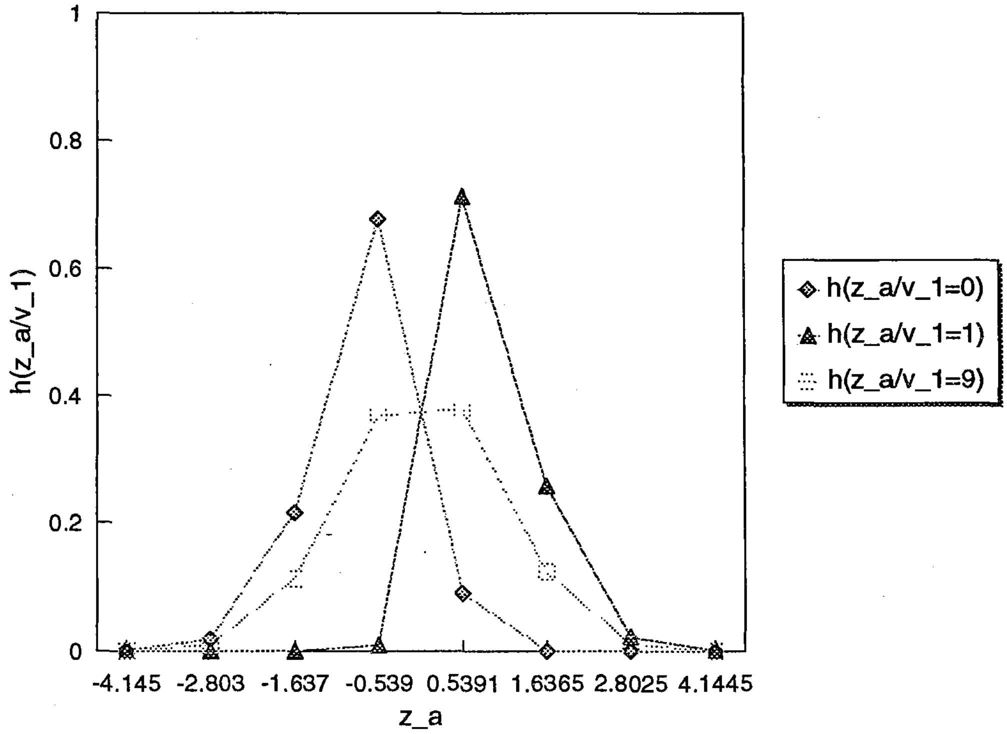

Figure 5.10: Item GAYTEASC, posterior probabilities

The scorings of individuals on the attitude scale based on their whole response pattern is given in Table (5.18). We see that someone who has not responded to the fourth item scores higher than someone who has the same answers to the rest of the items and responded to the fourth item as well.

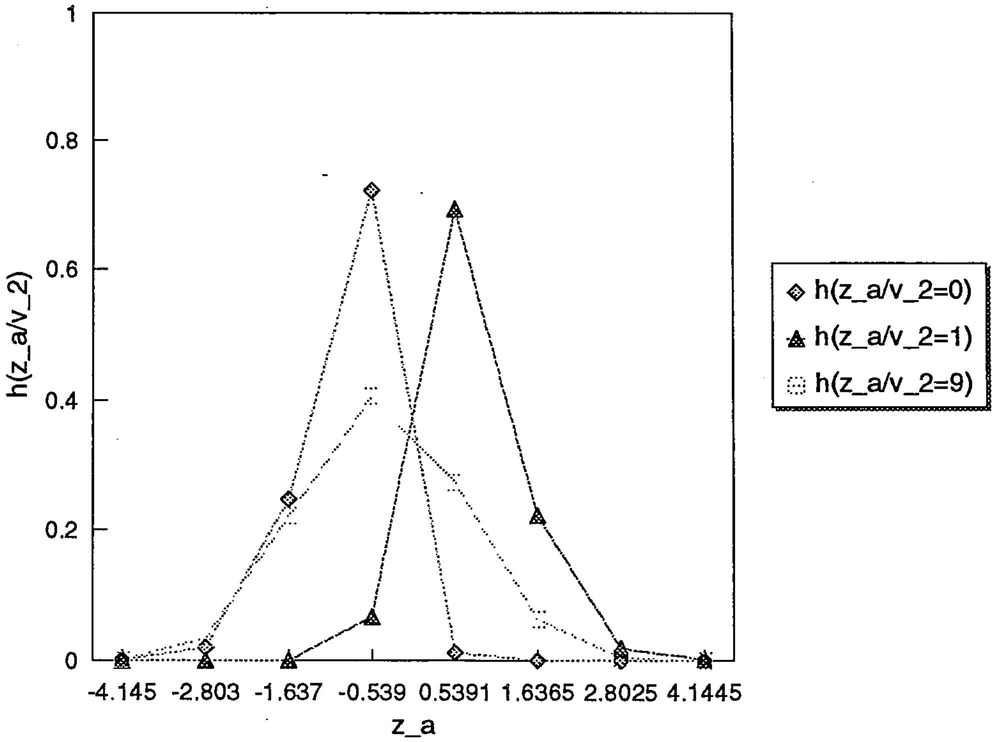

Figure 5.11: Item GAYTEAHE, posterior probabilities

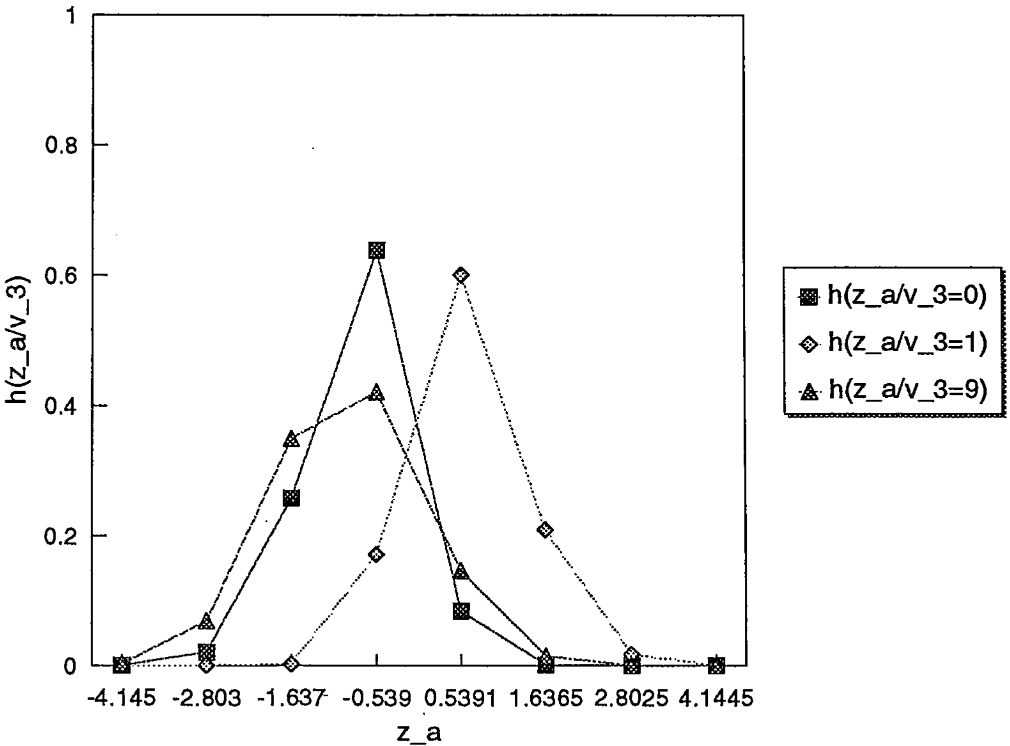

Figure 5.12: Item GAYPUB, posterior probabilities

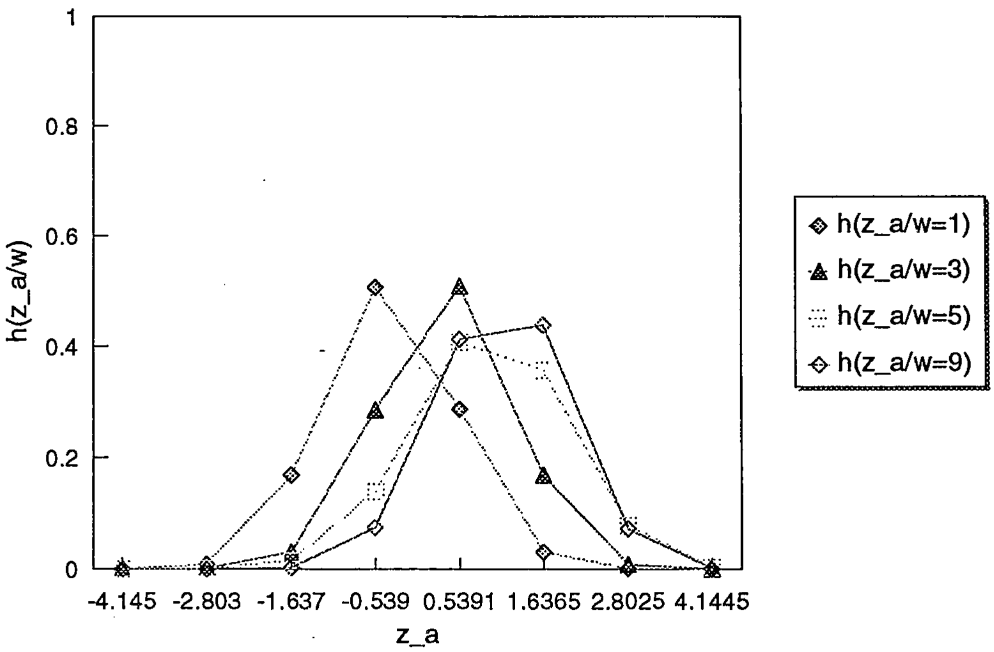

Figure 5.13: Item SAME SEX, posterior probabilities

Table 5.18: Posterior mean, sexual attitudes

$E(z_a \mid \mathbf{w}, \mathbf{v})$	responses	$E(z_a \mid \mathbf{w}, \mathbf{v})$	responses
-1.51	0991	0.00	0105
-1.20	0091	0.25	0111
-1.09	0901	0.26	1011
-0.92	0001	0.31	9191
-0.91	0902	0.33	9911
-0.81	0099	0.36	0112
-0.80	9001	0.36	1012
-0.78	0002	0.43	0113
-0.75	9991	0.47	0114
-0.68	0003	0.50	1015
-0.66	9901	0.53	1101
-0.62	0004	0.54	1102
-0.59	0005	0.55	1013
-0.55	0011	0.57	1104
-0.54	0012	0.57	1911
-0.53	0013	0.59	1105
-0.52	0014	0.62	1912
-0.50	0015	0.63	9111
-0.47	0911	0.67	1111
-0.46	9992	0.68	9913
-0.44	0101	0.76	1112
-0.37	0102	0.83	1195
-0.35	9903	0.88	9113
-0.34	0191	0.90	1113
-0.28	0103	1.08	9919
-0.23	0192	1.08	1114
-0.17	9909	1.32	1115
0.00	9101	1.35	1119

Chapter 6

Generalized latent trait models

6.1 Introduction

In this chapter we discuss the issue of generalizing the latent trait model for mixed manifest variables for types of distributions other than the Bernoulli and the Normal. The aim is to set up a general model framework from which manifest variables with different distributions in the exponential family can be analyzed with a latent trait model. A unified maximum likelihood method for estimating the parameters of the generalized latent trait model will be presented.

It will be shown that the latent trait model for mixed variables already developed in Chapter 2 can be generalized for other types of distributions and that all these different models share common characteristics and so a common method can be used for estimating model parameters.

In addition to the estimation of the latent trait model general results for the scoring methods (component score, posterior mean) will be presented.

In statistical theory generalized linear models (GLIM) were introduced by Nelder and Wedderburn (1972) and a systematic discussion of them can be found in McCullagh and Nelder (1989). The GLIM include as special cases, linear regression models with Normal, Poisson or Binomial errors and log-linear models. In all these models the explanatory variables are observed variables. In psychometric theory a similar generalization can be done in which the explanatory variables are latent (unobserved) variables.

Mellenbergh (1992) discusses the issue of putting the item response theory in a general framework. He refers to a number of different item formats such as dichotomous, polytomous, ordered polytomous and continuous items. As he noticed the latent variable models of these item formats can be described by a general model (GLIM) in which a monotone function of the expected response to an manifest item can be expressed as a linear function of latent variables and manifest explanatory variables. However, he does not discuss the possibility of having several types of distributions and he does not go into the problem of estimating the parameters of the generalized item response model.

In this chapter an attempt is made for putting in a general framework the latent trait model with mixed manifest variables.

6.2 Generalized linear models

A generalized linear model consists of three components:

1. The random component in which the random response variables, $(x_1, \cdots, x_p)$ have distributions from the exponential family, (such as Binomial, Poisson, Multinomial, Normal, Gamma).

2. The systematic component in which covariates, here the latent variables, $z_1, z_2, \cdots, z_q$ produce a linear predictor η:

$$\eta_i = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j, \qquad i = 1, \cdots, p$$

3. The link between the random and the systematic components:

$$\eta_i = g_i(\mu_i)$$

where $g_i(.)$ is called the link function which can be any monotonic differentiable function and different for different manifest variables.

Let $(x_1, x_2, \cdots, x_p)$ denote a vector of p manifest variables where each variable has a distribution in the exponential family taking the form:

$$f(x_i) = \exp\{\frac{x_i\theta_i - b_i(\theta_i)}{\alpha(\phi_i)} + c_i(x_i, \phi)\} \qquad i = 1, \cdots, p \tag{6.1}$$

where θ is called a canonical parameter and $\alpha(\phi)$, $b(\theta)$ and $c(x, \phi)$ are specific functions taking a different form depending on the distribution of the response variable x_i. More specifically $\alpha(\phi_i)$ is a scale parameter, taking commonly the form ϕ/w where w are known weights that may vary from observation to observation.

The mean and variance of the variable X can be derived from the relations based on the loglikelihood function $l(\theta, \phi; x) = \log f(x; \theta, \phi)$:

$$E(\frac{\partial l}{\partial \theta}) = 0$$

and

$$E(\frac{\partial^2 l}{\partial \theta^2}) + E(\frac{\partial l}{\partial \theta})^2 = 0$$

From these two equations we get that:

$$E(X) = b'(\theta)$$

and

$$Var(X) = b''(\theta)\alpha(\phi)$$

The variance of X is called the variance function.

Now we are going to identify for different types of responses the three components of the generalized model and the form of the specific functions given above. We will illustrate the models with one latent variable.

6.2.1 Binary responses

Let x_i take values 0 and 1, $(i = 1, \cdots, p)$. Suppose that the p manifest binary variables have Bernoulli distributions with expected value $\pi_i(z)$. The link function is defined to be the logit, i.e.:

$$g(\pi_i(z)) = \theta_i(z) = \text{logit}\pi_i(z) = \log(\frac{\pi_i(z)}{1 - \pi_i(z)}) = \alpha_{i0} + \alpha_{i1}z$$

where

$$\pi_i(z) = e^{\theta_i(z)}/(1 + e^{\theta_i(z)})$$

$$b_i(\theta_i(z)) = \log(1 + e^{\theta_i(z)})$$

$$\alpha(\phi_i) = 1$$

$$g(x_i \mid z) = \pi_i(z)^{x_i}(1 - \pi_i(z))^{1-x_i}$$

6.2.2 Normal distribution

let X_i be a normally distributed variable with mean μ_i and variance σ_i^2. The link function of the conditional distribution $x \mid z$ is the identity:

$$g(\mu_i) = \theta_i(z) = \mu_i(\theta) = \alpha_{i0} + \alpha_{i1}z = \mu_i + \lambda_{i1}z$$

Where $\mu_i + \lambda_{i1}z$ is a standard notation in the literature for the normal factor analysis model.

Also,

$$b_i(\theta_i(z)) = [\theta_i(z)]^2/2$$

$$\phi = \sigma^2$$

$$g(x_i \mid z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x_i - \mu_i - \lambda_{i1}z)^2\}$$

6.2.3 Gamma distribution

Let X_i distributed as a Gamma function. The link function is the reciprocal:

$$g(\mu_i) = \theta_i(z) = -\frac{1}{\gamma_i(z)} = \alpha_{i0} + \alpha_{i1} z$$

where $\gamma_i(z) = -\frac{1}{\theta_i(z)}$

$$b_i(\theta_i(z)) = -\log(-\theta_i(z)) = -\log(\frac{1}{\gamma_i(z)})$$

$$\phi = \frac{1}{\nu}$$

$$c_i(x;\phi) = \nu\log(\nu x) - \log x - \log\Gamma(\nu)$$

Hence,

$$\begin{aligned} g(x_i \mid z) &= \exp\{[-\frac{1}{\gamma_i}x_i + \log\frac{1}{\gamma_i}]/(1/\nu) + \nu\log(\nu x) - \log x - \log\Gamma(\nu)\} \\ &= \exp\{-\frac{x_i\nu}{\gamma_i}\}(\frac{1}{\gamma_i})^\nu(\nu x_i)^\nu x_i^{-1}\frac{1}{\Gamma(\nu)} \\ &= \frac{\exp\{-\frac{\nu}{\gamma_i}x_i\}x_i^{\nu-1}}{(\frac{\gamma_i}{\nu})^\nu\Gamma(\nu)} \end{aligned}$$

The shape parameter for the Gamma distribution is here $\nu = \frac{1}{\phi}$ or $\frac{w_i}{\phi}$ where w_i are prior weights and the scale parameter is $\frac{\gamma_i}{\nu} = \gamma_i\phi$ or $\frac{\gamma_i\phi}{w_i}$.

6.3 Estimation

The estimation of the parameters is based on the maximization of the joint distribution of the manifest variables. In this formulation of the model we allow the manifest variables to take any form from the exponential family.

Under the assumption of conditional independence the joint distribution of the manifest variables is:

$$f(\mathbf{x}) = \int_{-\infty}^{+\infty} g(\mathbf{x} \mid z)h(z)dz$$

$$= \int_{-\infty}^{+\infty} [\prod_{i=1}^{p} g_i(x_i \mid z)]h(z)dz$$

There is no constraint that the $g(x_i \mid z)$ for all the p items must be of the same type. Here, $g(x_i \mid z)$ can be any distribution from the exponential family.

For a random sample of size n the loglikelihood is written as:

$$\begin{aligned} L &= \sum_{h=1}^{n} \log f(\mathbf{x}_h) \\ &= \sum_{h=1}^{n} \log \int_{-\infty}^{+\infty} g(\mathbf{x}_h \mid z)h(z)dz \end{aligned} \tag{6.2}$$

The integral in equation (6.2) is approximated by Gauss-Hermite quadrature nodes and the loglikelihood to be maximized is written as:

$$L = \sum_{h=1}^{n} \log \sum_{t=1}^{k} [\prod_{i=1}^{p} \exp\{\frac{x_i\theta_i(z_t)}{\alpha(\phi_i)} - \frac{b_i(\theta_i(z_t))}{\alpha(\phi_i)} + c_i(\phi_i, x_i)\}]h(z_t) \tag{6.3}$$

The unknown parameters are in $\theta_i(z_t)$ and in $\alpha(\phi_i)$. Hence we have to differentiate the loglikelihood given in equation (6.3) with respect to the $\theta_i(z_t)$ and $\alpha(\phi_i)$ in order to obtain maximum likelihood estimates for the parameters, α_{i0} and α_{i1} and the scale parameter.

Finding partial derivatives, we have

$$\begin{aligned} \frac{\partial L}{\partial \alpha_{il}} &= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \frac{\partial f(\mathbf{x}_h)}{\partial \alpha_{il}} \\ &= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \sum_{t=1}^{\nu} g(\mathbf{x}_h \mid z_t)h(z_t) \frac{\partial}{\partial \alpha_{il}} [\frac{x_{ih}\theta_i(z_t)}{\alpha(\phi_i)} - \frac{b_i(\theta_i(z_t))}{\alpha(\phi_i)}] \end{aligned} \tag{6.4}$$

By interchanging the summation in equation (6.4) we get:

$$\frac{\partial L}{\partial \alpha_{il}} = \sum_{t=1}^{k} h(z_t)[\sum_{h=1}^{n} x_{ih} \frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)\alpha(\phi_i)} \frac{\partial \theta_i(z_t)}{\partial \alpha_{il}} - \sum_{h=1}^{n} \frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)\alpha(\phi_i)} \frac{\partial b_i(\theta_i(z_t))}{\partial \alpha_{il}}]$$

$$= \sum_{t=1}^{k}[r_{1it}\theta_i'(z_t) - N_t b_i'(\theta_i(z_t))]/\alpha(\phi_i) \tag{6.5}$$

where,

$$\theta_i'(z_t) = \frac{\partial\theta_i(z_t)}{\partial\alpha_{il}}$$

$$b_i'(\theta_i(z_t)) = \frac{\partial b_i(\theta_i(z_t))}{\partial\alpha_{il}}$$

$$\begin{aligned} r_{1it} &= h(z_t)\sum_{h=1}^{n} x_{ih} g(\mathbf{x}_h \mid z_t)/f(\mathbf{x}_h) \\ &= \sum_{h=1}^{n} x_{ih} h(z_t \mid \mathbf{x}_h) \end{aligned} \tag{6.6}$$

and

$$\begin{aligned} N_t &= h(z_t)\sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t)/f(\mathbf{x}_h) \\ &= \sum_{h=1}^{n} h(z_t \mid \mathbf{x}_h). \end{aligned} \tag{6.7}$$

Setting the partial derivatives equal to zero, (equation 6.5), we get:

$$\frac{\partial L}{\partial\alpha_{il}} = \sum_{t=1}^{k}[r_{1it}\theta_i'(z_t) - N_t b_i'(\theta_i(z_t))] = 0, \tag{6.8}$$

where the $b_i'(\theta_i(z_t))$ becomes:

Binary items: $b_i'(\theta_i(z_t)) = z_t^l \pi_i(z_t), \quad l = 0,1$

Normal continuous items: $b_i'(\theta_i(z_t)) = z_t^l(\mu_i + \lambda_{i1} z_t), \quad l = 0,1$

Gamma continuous items: $b_i'(\theta_i(z_t)) = z_t^l(-\frac{1}{\alpha_{i0}+\alpha_{i1} z_t}), \quad l = 0,1$

By replacing the above results into equation (6.8) we get:

For binary items:

$$\frac{\partial L}{\partial \hat{\alpha}_{il}} = \sum_{t=1}^{k} z_t^l [r_{1it} - N_t \pi_i(z_t)] = 0, \quad l = 0, 1 \tag{6.9}$$

For Normal continuous items:

$$\frac{\partial L}{\partial \hat{\alpha}_{il}} = \sum_{t=1}^{k} z_t^l [r_{1it} - N_t(\hat{\mu}_i + \hat{\lambda}_{i1} z_t)] = 0, \quad l = 0, 1 \tag{6.10}$$

where $\hat{\alpha}_{i0} = \hat{\mu}_i$ and $\hat{\alpha}_{i1} = \hat{\lambda}_{i1}$

For Gamma continuous items:

$$\frac{\partial L}{\partial \hat{\alpha}_{il}} = \sum_{t=1}^{k} z_t^l [r_{1it} + \frac{N_t}{\hat{\alpha}_{i0} + \hat{\alpha}_{i1} z_t}] = 0 \quad l = 0, 1 \tag{6.11}$$

The maximum likelihood equations for the binary and the Normal continuous items (6.9) and (6.10) respectively are the same as the ones obtained in Chapter 2. By formulating the model in this general way it is noticed that the derivatives of the loglikelihood respect to the unknown parameters can be very easily obtained for any type of distribution from the exponential family and the only information we need is the first derivative of the specific function $b_i(\theta_i(z_t))$.

For the Normal continuous items we get explicit formulae for the estimated parameters $\hat{\mu}_i$ and $\hat{\lambda}_{i1}$. For the binary and the Gamma continuous items the ML equations are non-linear equations respect to the parameters. The non-linear equations can be solved using a Newton-Raphson iterative scheme. The updating equation of the Newton-Raphson iterative solution is given by:

$$\hat{\beta}_{r+1} = \hat{\beta}_r - H^{-1}(\hat{\beta}_r) u(\hat{\beta}_r) \tag{6.12}$$

Where $\hat{\beta}$ denotes the vector with the unknown parameters, $H^{-1}(\hat{\beta}_r)$ is called Hessian matrix and contains the second derivatives of the loglikelihood respect to the unknown parameters and $u(\hat{\beta}_r)$ contains the first derivatives of the loglikelihood respect to the unknown parameters and r denotes the number of iteration.

A general form of the second derivatives can be found. The derivation is shown in Appendix D.

$$\frac{\partial^2 L}{\partial \alpha_{il}^2} = -\sum_{t=1}^{k} z_t^l N_t b_i''(\theta_i(z_t))/\alpha(\phi_i) \qquad l = 0,1 \tag{6.13}$$

$$\frac{\partial^2 L}{\partial \alpha_{i0}\partial \alpha_{i1}} = -\sum_{t=1}^{k} N_t b_i''(\theta_i(z_t))/\alpha(\phi_i) \tag{6.14}$$

Where $b_i''(\theta_i(z_t))$ denotes the second derivative respect to α_{il}. As it be seen from equations (6.13) and (6.14) the second derivatives depend on the second derivative of the specific function $b_i(\theta_i(z_t))$. This can be calculated for the different types of distributions.

For binary items:

$$\frac{\partial^2 b_i(\theta_i(z_t))}{\partial \alpha_{i0}^2} = \pi_i(z_t)(1 - \pi_i(z_t)) \tag{6.15}$$

$$\frac{\partial^2 b_i(\theta_i(z_t))}{\partial \alpha_{i1}^2} = z_t^2 \pi_i(z_t)(1 - \pi_i(z_t)) \tag{6.16}$$

For Normal continuous items:

$$\frac{\partial^2 b_i(\theta_i(z_t))}{\partial \mu_i^2} = 1 \tag{6.17}$$

$$\frac{\partial^2 b_i(\theta_i(z_t))}{\partial \lambda_{i1}^2} = z_t \tag{6.18}$$

For Gamma continuous items:

$$\frac{\partial^2 b_i(\theta_i(z_t))}{\partial \alpha_{i0}^2} = \frac{1}{(\alpha_{i0} + \alpha_{i1} z_t)^2} \tag{6.19}$$

$$\frac{\partial^2 b_i(\theta_i(z_t))}{\partial \alpha_{i1}^2} = \frac{z_t}{(\alpha_{i0} + \alpha_{i1} z_t)^2} \tag{6.20}$$

Now it remains to estimate the scale parameter ϕ. By differentiating the loglikelihood respect to the scale parameter ϕ we have:

$$\begin{aligned} \frac{\partial L}{\partial \phi_i} &= \sum_{h=1}^{n} \frac{1}{f(\mathbf{x}_h)} \sum_{t=1}^{k} h(z_t) g(\mathbf{x}_h \mid z_t) * \{ -\frac{x_{ih}\theta_i(z_t) - b_i(\theta_i(z_t))}{[\alpha(\phi_i)]^2} [\alpha(\phi_i)]' + \\ & c_i'(\phi_i, x_{ih}) \} \end{aligned} \tag{6.21}$$

By interchanging the summation in equation (6.21) and setting it equal to zero we have:

$$
\begin{aligned}
\sum_{t=1}^{k} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)} [b_i(\theta_i(z_t)) - x_{ih}\theta_i(z_t)][\alpha(\hat{\phi}_i)]' + \\
\sum_{t=1}^{k} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)} [c_i'(\phi_i, x_{ih}) * [\alpha(\hat{\phi}_i)]^2] = 0 \\
\sum_{t=1}^{k} [b_i(\theta_i(z_t))N_t - r_{1it}\theta_i(z_t)] * [\alpha(\hat{\phi}_i)]' + \\
\sum_{t=1}^{k} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)} c_i'(\phi_i, x_{ih}) * [\alpha(\hat{\phi}_i)]^2 = 0
\end{aligned}
\tag{6.22}
$$

By solving equation (6.22) respect to $[\alpha(\phi_i)]^2$ we have:

$$
[\alpha(\hat{\phi}_i)]^2] = \frac{\sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t]}{\sum_{t=1}^{k} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)} * c_i'(\hat{\phi}_i, x_{ih})} \tag{6.23}
$$

Equation (6.23) is true if $\alpha'(\phi_i) = 1$. The function $c_i'(\hat{\phi}_i, x_{ih})$ does depend on ϕ_i and so we do not get explicit form for ϕ_i.

More specifically for the different type of distributions, we have that for the Bernoulli distribution the scale parameter $\phi = 1$. For the Normal continuous items the form of $c_i(\phi_i, x_i)$ is given by:

$$
c_i(\phi_i, x_i) = -\frac{1}{2}(\frac{x_i^2}{\phi} + \log(2\pi\phi)) \tag{6.24}
$$

and

$$
c_i'(\phi_i, x_i) = \frac{x^2}{\phi^2} - \frac{1}{\phi} \tag{6.25}
$$

By replacing (6.25) into equation (6.23) we get:

$$
\hat{\phi}_i = \hat{\Psi}_i = \frac{1}{\sum_{t=1}^{k} N_t} \sum_{t=1}^{k} [r_{2it} - 2\hat{\mu}_i r_{1it} - 2\hat{\lambda}_{i1} z_t r_{1it} + (\hat{\mu}_i + \hat{\lambda}_{i1} z_t)^2 N_t] \tag{6.26}
$$

where,

$$
\begin{aligned}
r_{1it} &= \sum_{h=1}^{n} x_{ih} h(z_t \mid \mathbf{x}_h) \\
r_{2it} &= \sum_{h=1}^{n} x_{ih}^2 h(z_t \mid \mathbf{x}_h) \\
N_t &= \sum_{h=1}^{n} h(z_t \mid \mathbf{x}_h).
\end{aligned} \tag{6.27}
$$

For Gamma continuous variables the form of the $c_i(\phi_i, x_i)$ is given by:

$$c_i(\phi_i, x_i) = \nu \log \nu x - \log x - log\Gamma(\nu) \tag{6.28}$$

where $\phi = \frac{1}{\nu}$ or $\nu = \frac{1}{\phi}$. The first derivative of the function $c_i(\phi_i, x_i)$ required by formula (6.23) is:

$$
\begin{aligned}
c_i'(\phi_i, x_i) &= [\phi^{-1} \log \phi^{-1} x_i \log x_i - \log \Gamma(\phi^{-1})]' \\
&= -\frac{1}{\phi^2} \log \frac{x_i}{\phi} + \frac{1}{\phi} \frac{\phi}{x_i} (\frac{x_i}{\phi})' - \frac{1}{\Gamma(\phi^{-1})} [\Gamma(\phi^{-1})]'(\phi^{-1})' \\
&= -\frac{1}{\phi^2} \log \frac{x_i}{\phi} - \frac{1}{\phi^2} + \frac{1}{\phi^2} \frac{1}{\Gamma(\phi^{-1})} [\Gamma(\phi^{-1})]'
\end{aligned} \tag{6.29}
$$

From equation 6.23 we get:

$$
\begin{aligned}
\phi^2 &= \frac{\sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t]}{\sum_{t=1}^{k} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)} [-\frac{1}{\phi^2}[\log \frac{x_i}{\phi} + 1 - \frac{\Gamma(\phi^{-1})'}{\Gamma(\phi^{-1})}]]} \\
\Longleftrightarrow & \sum_{t=1}^{k} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)} [-\log \frac{x_h}{\phi} - 1 + \frac{\Gamma(\phi^{-1})'}{\Gamma(\phi^{-1})}] \\
&= \sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t] \\
\Longleftrightarrow & -\sum_{t=1}^{k} r_{3it} + \log \phi \sum_{t=1}^{k} N_t - \sum_{t=1}^{k} N_t + \sum_{t=1}^{k} h(z_t) \sum_{h=1}^{n} \frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)} \frac{\Gamma(\phi^{-1})'}{\Gamma(\phi^{-1})} \\
&= \sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t] \\
\Longleftrightarrow & \sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t] + \sum_{t=1}^{k} r_{3it} + \sum_{t=1}^{k} N_t
\end{aligned}
$$

$$
\begin{aligned}
&= \log\phi\sum_{t=1}^{k} N_t + \sum_{t=1}^{k} h(z_t)\sum_{h=1}^{n}\frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)}\frac{\Gamma(\phi^{-1})'}{\Gamma(\phi^{-1})} \\
&\iff \sum_{t=1}^{k}[r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t + r_{3it} + N_t] \\
&= \log\phi\sum_{t=1}^{k} N_t + \sum_{t=1}^{-k} h(z_t)\sum_{h=1}^{n}\frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)}\frac{\Gamma(\phi^{-1})'}{\Gamma(\phi^{-1})} \qquad (6.30)
\end{aligned}
$$

where,

$$
r_{3it} = h(z_t)\sum_{h=1}^{n}\log x_{ih} g(\mathbf{x}_h \mid z_t)/f(\mathbf{x}_h) \qquad (6.31)
$$

But $\Gamma(\phi^{-1})$ does not depend on x_h and so equation (6.30) becomes:

$$
\sum_{t=1}^{k}[r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t + r_{3it} + N_t] = [\log\phi + \frac{\Gamma(\phi^{-1})'}{\Gamma(\phi^{-1})}]\sum_{t=1}^{k} N_t \qquad (6.32)
$$

Before we proceed with equation (6.32) let investigate a Gamma model with both the response (x) and the explanatory (z) variables observed. The loglikelihood function of a random sample of size n is written:

$$
L = \sum_{h=1}^{n}\log\frac{1}{\Gamma(\nu)}(\frac{\nu x_h}{\mu})^{\nu} exp(-\frac{\nu x_h}{\mu})\frac{1}{x_h} \qquad (6.33)
$$

By differentiating the loglikelihood in equation (6.33) respect to the scale parameter ν and set it up equal to zero we get:

$$
\begin{aligned}
\frac{\partial L}{\partial \hat{\nu}} &= 0 \\
&\iff n\{\log\hat{\nu} - \frac{\Gamma(\hat{\nu})'}{\Gamma(\hat{\nu})}\} = \sum_{h=1}^{n}\{-\log\frac{x_h}{\hat{\mu}} + \frac{x_h - \hat{\mu}}{\hat{\mu}}\} \qquad (6.34)
\end{aligned}
$$

where,

$$
2\sum_{h=1}^{n}\{-\log\frac{x_h}{\hat{\mu}} + \frac{x_h - \hat{\mu}}{\hat{\mu}}\} = D(\mathbf{x};\hat{\mu}) \qquad (6.35)
$$

where $D(\mathbf{x};\hat{\mu})$ is called the deviance. The deviance is a goodness-of-fit measure that is defined to be the logarithm of the ratio of two likelihoods. One is the loglikelihood achievable for an exact fit in which the observed data are equal to the predicted data from the model and the other is the loglikelihood for the model under investigation.

$$D(\mathrm{x};\hat{\mu}) = 2\log l(\mathrm{x},\phi;\mathrm{x}) - 2\log l(\mu,\phi;\mathrm{x}) \tag{6.36}$$

For the models from the exponential family in which we denote the canonical parameters under the full model and the model under investigation to be $\tilde{\theta} = \theta(x)$ and $\hat{\theta} = \theta(\hat{\mu})$ respectively, the scaled deviance is written as:

$$\sum_{h=1}^{n} 2\{x_h(\tilde{\theta} - \hat{\theta}) - b(\tilde{\theta}) + b(\hat{\theta})\}/(\phi/w_i) = D(\mathrm{x};\hat{\mu})/\phi \tag{6.37}$$

$D(\mathrm{x};\hat{\mu})$ is known as the deviance for the current model and is a function of the data only. The scaled deviance is defined as:

$$D^*(\mathrm{x};\hat{\mu}) = D(\mathrm{x};\hat{\mu})/\phi \tag{6.38}$$

and it expressed as a multiple of the dispersion parameter.

The forms of the deviances for the distributions of the exponential family are given in McCullagh and Nelder (1989), page 34. For binary responses the scaled deviance is:

$$D^*(\mathrm{x};\hat{\mu}) = 2\sum_{h=1}^{n}\{x_h \log\frac{x_h}{\hat{\mu}} + (1 - x_h)\log\frac{1 - x_h}{1 - \hat{\mu}}\}$$

For Normal continuous variables the scaled deviance is:

$$D^*(\mathrm{x};\hat{\mu}) = \sum_{h=1}^{n}(x_h - \hat{\mu})^2/\sigma^2$$

which is the residual sum of squares. For the Gamma continuous variables as we have already shown is:

$$D^*(\mathrm{x};\hat{\mu}) = 2\sum_{h=1}^{n}\{-\log\frac{x_h}{\hat{\mu}} + \frac{x_h - \hat{\mu}}{\hat{\mu}}\}$$

In the generalized linear models there is only one response variable x. In the latent variable models there is a vector of p response (manifest) variables $\mathbf{x}$. As a result we need a deviance for each manifest variable $x_i, i = 1, \cdots, p$.

The reason for introduction of the deviance is that we can use it to get an estimate of the scale parameter when we do not get an explicit solution from maximum likelihood estimation. Estimation of the scale parameter ϕ is not needed for the Bernoulli and the Poisson distribution because it is taken to be equal to one and for the Normal distribution we get an explicit form for the estimator of the scale parameter. Now for the Gamma distribution the scale parameter can be estimated if subroutines that calculate the digamma, gamma and trigamma functions are provided. Another way to estimate the scale parameter is to use the deviance. We are now going to investigate that.

The statistic $S = D/\phi$ can be computed from the data. Now under a reasonable model the statistic S has a χ^2 distribution with mean $E(S) = n - r$ where r are the number of independent linear parameters that are estimated. By making the assumption that under a reasonable model the S statistic will be close to its expected value we can estimate the scale parameter from:

$$\hat{\phi} = \frac{D}{n-r}$$

In Francis, Green, and Payne (1993) is mentioned that the estimate $\hat{\phi}$ is inconsistent as $n \rightarrow \infty$.

Let us see now how equation (6.32) is related to (6.35) which is the deviance in the generalized linear model for Gamma distribution. Equation (6.32) becomes:

$$\sum_{t=1}^{k}[r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t + r_{3it} + N_t] = [\log\phi + \frac{\Gamma(\phi^{-1})'}{\Gamma(\phi^{-1})}]\sum_{t=1}^{k} N_t \iff$$

$$-\sum_{t=1}^{k} N_t \frac{x_{ih}}{\hat{\mu}_i} + \sum_{t=1}^{k} \log N_t(\frac{1}{\hat{\mu}_i}) + \sum_{t=1}^{k} N_t \log x_{ih} + \sum_{t=1}^{k} N_t = [\log\phi + \frac{\Gamma(\phi^{-1})'}{\Gamma(\phi^{-1})}]\sum_{t=1}^{k} N_t \iff$$

$$-[\sum_{t=1}^{k} N_t \log\frac{x_{ih}}{\hat{\mu}_i} - \sum_{t=1}^{k} N_t \frac{(x_{ih} - \hat{\mu}_i)}{\hat{\mu}_i}] = \sum_{t=1}^{k} N_t[\log\phi^{-1} - \frac{\Gamma(\phi^{-1})'}{\Gamma(\phi^{-1})}] \quad (6.39)$$

By replacing $\phi = 1/\nu$ into equation (6.39) and multiple both sides by 2 we get:

$$-2[\sum_{t=1}^{k} N_t \log \frac{x_{ih}}{\hat{\mu}_i} - \sum_{t=1}^{k} N_t \frac{(x_{ih} - \hat{\mu}_i)}{\hat{\mu}_i}] = 2\sum_{t=1}^{k} N_t[\log \hat{\nu} - \frac{\Gamma(\hat{\nu})'}{\Gamma(\hat{\nu})}] \tag{6.40}$$

The right expression of equation (6.40) is the same with the right expression of equation (6.35). So the left hand side of equation (6.40) must be the deviance for the latent variable model with Gamma conditional distributions for the manifest variables.

An approximation which is used for equation 6.35 is:

$$\hat{\nu}^{-1} = \frac{\bar{D}(6+\bar{D})}{6+2\bar{D}} \tag{6.41}$$

where $\bar{D} = D(\mathrm{x}; \hat{\mu})/n$. The same approximation might be used for the latent variable models.

6.3.1 EM Algorithm

So far we have used the letter x to denote the manifest variables assuming that x can be either binary or metric. Let denote with v the binary items, with w the normal continuous variables and with u the Gamma continuous variables.

The maximization of the loglikelihood (equation 6.3) is done by an E-M algorithm. This is the same algorithm described in Chapter 2 (section 2.2.4) for the latent trait model with mixed manifest variables. The steps of the algorithm are defined as follows:

step 1 Choose initial estimates for the model parameters (α_{il} and the scale parameter).

step 2 Compute the values $r_{1it}(v)$, $r_{1it}(w)$, $r_{1it}(u)$, $r_{2it}(w^2)$, $r_{3it}(\log u)$ and N_t.

step 3 Obtain improved estimates for the parameters by solving the non linear maximum likelihood equations for Bernoulli and Gamma distributed variables and using the explicit equations for Normal distributed variables.

step 4 Return to step 2 and continue until convergence is attained.

6.4 Scoring methods for the generalized latent trait model

It is appropriate if possible to find a general framework for the scoring methods for the generalized latent trait model. If the latent trait model fits the data then we can summarize the information in a set of manifest variables by obtaining a score on the latent dimension. The work in this section is an extension of the scoring methods already discussed in Chapter 2 for the latent trait model with mixed variables (binary and Normal continuous variables). Here, we would try to derive a general formula for the component scores which can be used under any type of distribution or mixture of distributions in the exponential family.

For the time being we will assume that all x's are of the same type. The conditional distribution of the response pattern $\mathbf{x}$ given z is in the exponential family and it takes the form:

$$\begin{aligned} g(\mathbf{x} \mid z) &= \prod_{i=1}^{p} g(x_i \mid z) \\ &= \prod_{i=1}^{p} \exp\{\frac{x_i\theta_i(z) - b_i(\theta_i(z))}{\alpha_i(\phi_i)} + c_i(x_i, \phi)\} \end{aligned} \tag{6.42}$$

where $\theta_i(z) = \alpha_{i0} + \alpha_{i1}z$.

Equation (6.42) becomes:

$$\begin{aligned} g(\mathbf{x} \mid z) &= \exp[\sum_{i=1}^{p} \frac{\alpha_{i0}}{\alpha(\phi_i)}x_i + \sum_{i=1}^{p} \frac{\alpha_{i1}}{\alpha(\phi_i)}x_i z - \sum_{i=1}^{p} \frac{b_i(\theta_i(z))}{\alpha(\phi_i)} + \sum_{i=1}^{p} c_i(x_i, \phi)] \\ &= \exp[c_o(\mathbf{x}) + c_1(\mathbf{x})z] \exp[-\sum_{i=1}^{p} \frac{b_i(\theta_i(z))}{\alpha(\phi_i)} + \sum_{i=1}^{p} c_i(x_i, \phi)] \end{aligned} \tag{6.43}$$

where $c_o(\mathbf{x}) = \sum_{i=1}^{p} \frac{\alpha_{i0}}{\alpha(\phi_i)}x_i$ and $c_1(\mathbf{x}) = \sum_{i=1}^{p} \frac{\alpha_{i1}}{\alpha(\phi_i)}x_i$

The conditional distribution of zero responses to all items given the latent variable is:

$$g(\mathbf{0} \mid z) = \prod_{i=1}^{p} \exp\{-\frac{b_i(\theta_i(z))}{\alpha(\phi_i)} + c_i(\mathbf{0}, \phi)\} \tag{6.44}$$

From equations (6.43) and (6.44) we have:

$$\begin{aligned}
g(\mathbf{x} \mid z) &= \exp[c_o(\mathbf{x}) + c_1(\mathbf{x})z] \prod_{i=1}^{p} \exp[-\frac{b_i(\theta_i(z))}{\alpha(\phi_i)} + c_i(x_i, \phi)] \\
&= \exp[c_o(\mathbf{x}) + c_1(\mathbf{x})z] \prod_{i=1}^{p} \exp[-\frac{b_i(\theta_i(z))}{\alpha(\phi_i)} + c_i(x_i, \phi)] \frac{\exp(c_i(\mathbf{0}, \phi_i))}{\exp(c_i(\mathbf{0}, \phi_i))} \\
&= \exp[c_o(\mathbf{x}) + c_1(\mathbf{x})z] \prod_{i=1}^{p} \exp[-\frac{b_i(\theta_i(z))}{\alpha(\phi_i)} + c_i(\mathbf{0}, \phi)] \frac{\exp(c_i(x_i, \phi_i))}{\exp(c_i(\mathbf{0}, \phi_i))} \\
&= \exp[c_o(\mathbf{x}) + c_1(\mathbf{x})z] g(\mathbf{0} \mid z) \prod_{i=1}^{p} \frac{\exp(c_i(x_i, \phi_i))}{\exp(c_i(\mathbf{0}, \phi_i))} \qquad (6.45)
\end{aligned}$$

The joint probability of the manifest variables $(\mathbf{x})$ may be written as:

$$\begin{aligned}
f(\mathbf{x}) &= \int_{-\infty}^{+\infty} g(\mathbf{x} \mid z) h(z) dz \\
&= \int g(\mathbf{0} \mid z) \exp(c_o(\mathbf{x}) + c_1(\mathbf{x})z) \prod_{i=1}^{p} \frac{\exp(c_i(x_i, \phi_i))}{\exp(c_i(\mathbf{0}, \phi_i))} h(z) dz \\
&= (\exp(c_o(\mathbf{x}))) \prod_{i=1}^{p} \frac{\exp(c_i(x_i, \phi_i))}{\exp(c_i(\mathbf{0}, \phi_i))} f(\mathbf{0}) \int_{-\infty}^{+\infty} \frac{g(\mathbf{0} \mid z)}{f(\mathbf{0})} \exp(c_1(\mathbf{x})z) h(z) dz \\
&= (\exp(c_o(\mathbf{x}))) \prod_{i=1}^{p} \frac{\exp(c_i(x_i, \phi_i))}{\exp(c_i(\mathbf{0}, \phi_i))} f(\mathbf{0}) \int_{-\infty}^{+\infty} h(z \mid \mathbf{0}) \exp(c_1(\mathbf{x})z) dz \\
&= (\exp(c_o(\mathbf{x}))) \prod_{i=1}^{p} \frac{\exp(c_i(x_i, \phi_i))}{\exp(c_i(\mathbf{0}, \phi_i))} f(\mathbf{0}) M_{z|0}(c_1(\mathbf{x})(\mathbf{x}) \qquad (6.46)
\end{aligned}$$

where, $M_{z|0}$ is the moment generating function of the conditional distribution of the latent variable z given a zero response on all items.

Hence, the conditional distribution of z given the response pattern x is:

$$\begin{aligned}
h(z \mid \mathbf{x}) &= \frac{g(\mathbf{x} \mid z) h(z)}{f(\mathbf{x})} \\
&= \frac{g(\mathbf{0} \mid z) \exp(c_o(\mathbf{x}) + c_1(\mathbf{x})z) \prod_{i=1}^{p} \frac{\exp(c_i(x_i, \phi_i))}{\exp(c_i(\mathbf{0}, \phi_i))} h(z)}{\exp(c_o(\mathbf{x})) f(\mathbf{0}) \prod_{i=1}^{p} \frac{\exp(c_i(x_i, \phi_i))}{\exp(c_i(\mathbf{0}, \phi_i))} M_{z|0}(c_1(\mathbf{x}))} \\
&= \frac{\exp(c_1(\mathbf{x})z) g(\mathbf{0} \mid z) h(z)}{f(\mathbf{0}) M_{z|0}(c_1(\mathbf{x}))} \qquad (6.47)
\end{aligned}$$

In the case that more than one type of manifest variables are fitted the part which is influenced in equation (6.47) is that that depends on the manifest variables i.e. $c_1(\mathbf{x})$. For the case we have for example three different type of manifest variables, (v, w, u), equation (6.47) becomes:

$$h(z \mid \mathbf{x}) = \frac{\exp\{(c_1(\mathbf{v}) + c_1(\mathbf{w}) + c_1(\mathbf{u}))z\}g(\mathbf{0} \mid z)h(z)}{f(\mathbf{0})M_{z|0}(c_1(\mathbf{v}) + c_1(\mathbf{w}) + c_1(\mathbf{u}))} \tag{6.48}$$

From equation (6.48) we see that the component scores are:

For binary items: $c_1(\mathbf{v}) = \sum_i \alpha_{i1} v_i$ since $\alpha(\phi) = 1$

For Normal continuous items: $c_1(\mathbf{w}) = \sum_i \frac{\lambda_{i1}}{\Psi_{ii}} w_i$ since $\alpha(\phi_i) = \Psi_i$

For Gamma continuous items: $c_1(\mathbf{u}) = \sum_i \frac{\alpha_{i1}}{1/\nu} u_i$

Hence, the component score for each response pattern/individual of the model with variables (v, w, u) each of different type is:

$$\sum_i \alpha_{i1} v_i + \sum_i \frac{\lambda_{i1}}{\Psi i} w_i + \sum_i \frac{\alpha_{i1}}{1/\nu} u_i \tag{6.49}$$

From equation (6.48), the moment generating function of the conditional distribution of z given $\mathbf{x}$ is

$$\begin{aligned} M_{z|x}(t) &= \int_{-\infty}^{+\infty} \exp(tz) h(z \mid \mathbf{x}) dz \\ &= \int \exp(tz) \frac{\exp[(c_1(\mathbf{v}) + c_1(\mathbf{w}) + c_1(\mathbf{u}))z] h(z \mid 0)}{M_{z|0}(c_1(\mathbf{v}) + c_1(\mathbf{w}) + c_1(\mathbf{u}))} dz \\ &= \frac{M_{z|0}(c_1(\mathbf{v}) + c_1(\mathbf{w}) + c_1(\mathbf{u}) + t)}{M_{z|0}(c_1(\mathbf{v}) + c_1(\mathbf{w}) + c_1(\mathbf{u}))} \end{aligned} \tag{6.50}$$

In Chapter 2, section 2.3 the results of Knott and Albanese (1993) for the latent trait model with binary items extended for the latent trait model for mixed variables. The same results can be also applied here for the generalized latent trait model.

Result 1 If $K_{z|0}(t)$ is the cumulant generating function for the density of z given that all responses are zero, then

$$E(z \mid \mathbf{x}) = K'_{z|0}(c_1(\mathbf{v}) + c_1(\mathbf{w}) + c_1(\mathbf{u})) \tag{6.51}$$

and

$$Var(z \mid \mathbf{x}) = K''_{z|0}(c_1(\mathbf{v}) + c_1(\mathbf{w}) + c_1(\mathbf{u})) \tag{6.52}$$

where the prime and double prime indicate first and second derivatives of the cumulant generating function.

Result 2 $E(z \mid \mathbf{v}, \mathbf{w}, \mathbf{u})$ is a strictly increasing function of $(c_1(\mathbf{v})+c_1(\mathbf{w})+c_1(\mathbf{u}))$, if the variance of the conditional distribution of z given that all responses are zero has variance strictly greater than zero. This results has been discussed in Chapter 2.

Result 3 When the conditional distribution of z when all responses are zero is normal, then the conditional distribution of z for any set of responses is normal.

6.5 Summary

In this section we want to summarize the results we found for the generalized linear latent trait model for mixed variables. The results presented in the previous sections used only one latent variable. Here we assume more than one latent variables.

Let the responses $(x_1, x_2, \cdots, x_p)$ are of different type formats. The different type formats can be binary items denoted by (v), normally distributed continuous items denoted by (w), gamma distributed continuous items denoted by (u), poisson count items denoted by (u^*) or any other type that can be in the exponential family.

The generalized latent trait model is written as:

$$g_i(\mu_i) = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j \quad i = 1, \cdots, p \tag{6.53}$$

where z denotes the latent variables and $g_i(.)$ can be any monotonic differentiable function taking different forms for different items depending on their distribution assumed. The latent variables are assumed to have independent standard normal distributions.

The estimation of the model parameters α_{i0} and α_{ij} is based on the maximization of the loglikelihood of the joint distribution of the manifest variables which under the assumption of conditional independence is written:

$$f(\mathbf{x}) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(\mathbf{v} \mid \mathbf{z}) g(\mathbf{w} \mid \mathbf{z}) g(\mathbf{u} \mid \mathbf{z}) g(\mathbf{u}^* \mid \mathbf{z}) dz \tag{6.54}$$

where the loglikelihood for a random sample of size n is:

$$L = \sum_{h=1}^{n} \log f(\mathbf{x}_h)$$

The maximum likelihood equations respect to the unknown parameters are:

$$\frac{\partial L}{\partial \alpha_{il}} = \sum_{t=1}^{k} z_t^l [r_{1it} - N_t b_i'(\theta_i(z_t))] = 0 \quad l = 0, 1 \tag{6.55}$$

where r_{1it} takes a different form depending on the type format of the ith manifest variable. The only think required to be calculated is the first derivative of the function $b_i(\theta_i(z_t))$. Non-linear equations can be solved with an iterative algorithm such as the Newton Raphson.

For the estimation of the scale parameter either the ML estimate is used if an explicit solution exists or an estimate based on the deviance can be obtained.

The maximization of the loglikelihood is done via an E-M algorithm described in section 6.3.1.

Chapter 7

Contribution of the Research

The purpose of this chapter is to outline the main findings of this research, to discuss the usefulness of the development of the theory and its applications and to discuss possible extensions of the present research in the future. The first section of this chapter is an overview of the research developments, as they have been presented in the previous chapters, the second section is a discussion of the limitations of the current research and some proposals for future research and the last section is a conclusion of the theoretical and practical contribution of the research.

7.1 Overview

This thesis has dealt with the problem of fitting a latent variable model to a number of mixed observed variables with complete and incomplete data. The mixed observed variables can be either binary or metric (discrete and continuous). We are interested in the estimation of the model parameters but also emphasis has been given in scoring methods for allocating individuals into the latent dimension based on their response patterns, in the incorporation of missing values into the analysis and the application of these methods into real data sets.

In Chapter 1 a discussion of the literature in the area of latent variable models for mixed observed variables is given. Although our work in this thesis is concentrated in the analysis of mixed variables an overview of the methods for binary and metric observed variables is given since many of these results have been extended here for the mixed model. We discuss the two approaches for the estimation of latent variable models. These are the underlying variable approach that assumes that an underlying variable exists for each binary or categorical observed variable and carries on the analysis on the correlation matrix of the underlying variables. The second approach analyzes the data as they are and defines for each individual in the sample the probability of responding positively to a variable given the individual's position on the latent factor space.

Significant contributions based on the underlying variable approach have been made by Muthén, Arminger and Küster as well as Jöreskog and Sörbom but no work has been done so far on the second approach.

The second approach has been well explored in the case where the observed variables are either binary or metric, see Bartholomew (1987). Our work here is an extension of Bartholomew's work for mixed type variables. Our approach is based on Bartholomew (1987) sufficiency principle that looks for summary statistics based on the observed variables that could contain all the information about the latent variables.

In Chapter 2 a latent trait model (continuous latent variables) is developed for fitting a number p of mixed observed variables, binary, and metric variables that are normally distributed. The results presented are for fitting a latent trait model with q latent variables where q is much less than p. We assume that the conditional distribution of the observed variables given the vector of latent variables follows a Bernoulli distribution for the binary variables and a normal distribution for the metric variables. The model developed here analyzes the response patterns as they are in contrast with the underlying variable approach. A discussion on the comparison of the two approaches is given in Chapter 2, section 2.2.8.

Marginal maximum likelihood estimation is used for estimating the model parameters via an E-M algorithm. Standard errors for the parameter estimates are obtained based on the asymptotic theory for maximum likelihood estimation.

As far as concern the goodness-of-fit for the model no statistical criterion has been used. However we look at the fit of the model in the two- and three-way margins of the binary variables and the covariance matrix for the metric variables.

Model selection criteria are also discussed.

A standardized solution is proposed to be used for a more unified interpretation of the results. Orthogonal transformation of the maximum likelihood solution is shown to be possible.

Finally, scoring methods for allocating individuals into the latent space are presented based on the posterior mean of the latent variable given the response pattern of each individual and the component score.

In Chapter 3 a latent class model (discrete latent variables) is developed for fitting mixed observed variables, binary and metric variables that are normally distributed. A similar theory which developed for the latent trait model for mixed observed variables is set up here for latent class models. Theory that already existed for binary, metric and mixed variables is presented in the beginning of the chapter. The only contribution in the literature for mixed type variables is by Everitt and Merette which assumes underlying variables for the observed categorical variables. Their method involves, as it has been shown, heavy integrations. Our method does not require any integrations and for that reason it is more advantageous than the existing method. Standard errors for the estimated parameters are provided from asymptotic maximum likelihood theory. Allocation of individuals into the latent classes is discussed.

In Chapter 4 four data sets have been analyzed using the models which developed in Chapter 2 and 3. The four data sets vary in the number of observed variables and the number of response patterns to be analyzed. Two of the data sets come from the British Social Attitudes Survey of 1990 and 1991 and the other two are from an LSE cognitive laboratory experiment. We discuss for all the models we fit the interpretation of the model parameters and scoring methods for the individuals. For goodness-of-fit we looked at the two- and three-way margins for the binary part of the model and at the sample and the one obtained under the model correlation matrix for the continuous part of the model. Akaike's criterion has been used as a model selection criterion. We should note that the metric variables used here are all four- or five-point scale variables treated as interval scale variables.

In Chapter 5 the latent trait model for mixed observed variables is expanded so that missing values can be handled. We are interested in item nonresponse. The set up and the estimation of the model are an extension of the work of Albanese and Knott (1992) for binary items with missing values. Apart from estimating the model parameters we give a lot of emphasis to how we can use the model to derive information about attitude from non-expression and how we can score individuals on the attitude latent dimension based on their response pattern. Artificial examples have been used with Guttman scale, non-type scale and Likert scale observed variables to illustrate the model plus a real data set from the British Social Attitudes Survey.

In Chapter 6 the results presented in Chapter 2 are generalized to observed variables with conditional distributions in the exponential family. Our aim is to develop a general framework like GLIM that can handle any type of observed variables in the exponential family such as the Binomial, Poisson, Normal and Gamma distribution. In Chapter 6 we show that this general framework exists and that a common estimation method can be used to find estimates for the model parameters. We work with the general form of the exponential family. The results of Chapter 2 can be derived as special cases for Bernoulli and Normally distributed variables. Also the issue of scoring individuals on the latent factor space has been looked at and is put in the general framework.

7.2 Limitations and future research

In this section we discuss the limitations of our research and problems that have risen. These limitations suggest areas for further research.

There is no formal statistical test at the moment for testing the goodness-of-fit of the models developed here for mixed variables (latent trait and latent class). So far the goodness-of-fit has been looked at separately for the binary and the metric observed variables.

The E-M algorithm suggested here for the maximization of the loglikelihood function converges very slowly if the number of latent variables is greater than two. Possible acceleration routines should be looked at.

The thesis has developed a latent trait model and a latent class model for binary and metric observed variables that are assumed to have conditional on the latent variables Bernoulli and normal distributions respectively. This part of the theory has been supported by software (LATENT and CLASSMIX). In the thesis we have further discussed the generalization of the latent trait model for other types of distributions in the exponential family, (chapter 6). This is not yet supported by software.

The generalized theory could be further extended in the future to cover also the latent class model.

Finally, it will be interesting to extend the model for categorical nominal and ordinal observed variables. This will make the model even more generally applicable.

7.3 Conclusion

In this section we would like to summarize the results of our research in order to show the contribution of this research in the area of latent variable models.

A fourfold classification is used in Bartholomew 1987 to classify the techniques available in the area of latent variable models. This is shown in the first two columns of the table given below. We discussed in Chapter 1 how Bartholomew 1987 presented a unified approach for fitting latent variable models for either categorical or continuous observed variables and for discrete or continuous latent variables. Our research is an extension of Bartholomew's work for mixed observed variables.

Our methodology developed latent variable methods for the analysis of mixed observed variables with complete and incomplete data. The main contribution of this thesis lies in the last column of the table below. After we have developed the theory for the mixed case many different issues were looked at and these are discussed briefly in the first section of this chapter.

		Manifest variables		
		Metrical	Categorical	Mixed
Latent variables	Metrical	factor analysis	latent trait analysis	latent trait analysis
	Categorical	latent profile analysis	latent class analysis	latent class analysis

Apart from the development of the methodology for handling mixed observed variables with latent variables emphasis is given in the application of this methodology in real problems. Data sets with mixed variables have been analyzed so that the advantages of the theory developed will be apparent. Two software programs have been written to fit the models developed.

Social scientists such as psychometricians and sociometricians use latent analysis to describe attitude relations. Economists and policy makers also use latent analysis to construct a wealth indicator from observed indicators such as income, expenditure, etc.. In all these disciplines it is common that the observed variables to be used have different level of measurements, such as binary and metric (discrete and continuous). By extending the existing theory of latent variable models for mixed data we have made an essential contribution in this area.

Appendix A

Program LATENT

The computer program LATENT fits a one and two factor latent trait model to mixed observed variables with complete and incomplete data. The program is written in FORTRAN 77.

The theoretical development for the latent trait model for mixed observed variables with complete data is presented in Chapter 2 and for the incomplete data in Chapter 5.

The program provides for both models parameter estimates, standard errors of the parameter estimates, scoring methods that based on the component score and the posterior mean and finally for a goodness-of-fit measure computes the observed and expected under the model first, second and third order margins.

Description of the program's input

The three lines below are the input lines required to be read by the program. These three lines together with the data set are saved in a file called LAT.INP.

TITLE
N NPD NPC NQ
MODEL NFAC INPUT FREQ DISPLAY MTER LOUT8 ERRC
Individual's response patterns are displayed here.

N denotes the number of response patterns to be analyzed.

NPD denotes the number of binary observed variables.

NPC denotes the number of metric observed variables.

NQ denotes the number of quadrature points to be used (8,16,24,32,48).

MODEL for complete data it takes the value 1 and for incomplete the value 3.

NFAC denotes the number of latent variables to be fitted (1 or 2).

INPUT takes the value 0 if the initial parameter values are set in the program and 1 if they have to be read by an external file called LAT3.INP.

FREQ takes the value 0 if raw data are going to be read and the value 1 if the frequency of the response patterns is going to be read.

DISPLAY takes the value 1 if a display of the frequency distribution is needed and 0 otherwise.

MTER denotes the maximum number of iterations.

LOUT8 takes the value 0 if a file with the final parameter estimates is needed to be saved and 1 otherwise.

ERRC denotes the convergence tolerance for the EM algorithm.

Brief description of the program's subroutines

- **EM1** This subroutine does the maximization step of the E-M algorithm. New estimates of the model parameters are obtained for the discrete and the continuous part of the model. The convergence of the E-M algorithm is checked.

- **EM2** Same as EM1 but for the two factor model.

- **PHILIKD** Computes the response function for each item $\pi_i(\mathbf{z})$ and the conditional distribution of the binary variables given the vector of the latent variables for each individual h, $g(\mathbf{v}_h \mid \mathbf{z})$

- **PHILIKC** Computes the conditional distribution of the metric variables given the vector of latent variables for each individual h, $g(\mathbf{v}_h \mid \mathbf{z})$.

- **PHILIK** Computes the conditionsl distribution of the binary and the metric variables given the vector of latent variables for each individual h, $g(\mathbf{v}_h, \mathbf{w}_h \mid \mathbf{z})$ and the loglikelihood value.

- **ENER** This subroutine does the expectation step of the E-M algorithm.
- **VARIANCE** Computes the variance-covariance matrix of the parameter estimates.
- **POSMEAN** Computes the posterior mean, $E(z \mid \mathbf{v}_h, \mathbf{v}_h)$, for each response pattern/ individual.
- **COMPONENT** Computes the component score for each response pattern/individual.
- **MARGIN** Computes the observed and expected one- two- and three-way margins.
- **CODING** Generates the pseudo variables required for the fit of model 2 (missing cases).
- **POSTERIOR** Computes the posterior values $h(z_a \mid v_i = 0)$,$h(z_a \mid v_i = 1)$, and $h(z_a \mid v_i = 9)$ for the binary variables and $h(z_a \mid W_i = w_i)$ for the metric variables required in the posterior analysis presented in Chapter 5 (section 5.2.3).

Appendix B

Program CLASSMIX

The computer program CLASSMIX fits a latent class model to mixed observed variables. The program is written in FORTRAN 77.

The theoretical development for the latent class model for mixed observed variables is presented in Chapter 3.

The program provides estimates of the model parameters, allocation of individuals into classes based on their response patterns and observed and expected frequencies for each response pattern (only for the binary items).

Description of the program's input

The three lines below are the input lines required to be read by the program. These three lines together with the data set are saved in a file called CLASS.INP.

TITLE
N NPD NPC
NC INPUT FREQ DISPLAY MTER LOUT8 ERRC
Individual's response patterns are displayed here.

N denotes the number of response patterns to be analyzed,
NPD denotes the number of binary observed variables
NPC denotes the number of metric observed variables
NC denotes the number of classes to be fitted $(1, \cdots, k)$

INPUT takes the value 0 if the initial parameter values are set in the program and 1 if they have to be read by an external file called CLASS3.INP.

FREQ takes the value 0 if individual response patterns are going to be read and the value 1 if the frequency of the response patterns is going to be read.

DISPLAY takes the value 1 if a display of the frequency distribution is needed and 0 otherwise.

MTER denotes the maximum number of iterations.

LOUT8 takes the value 0 if a file with the final parameter estimates is needed to be saved and 1 otherwise.

ERRC denotes the convergence tolerance for the E-M algorithm.

Brief description of the program's subroutines

- **EM** This subroutine's task is to control the expectation and maximization step of the E-M algorithm by actually calling two other subroutines one for each step. The convergence of the E-M algorithm is checked.
- **DISTRIB** This subroutine does the E-step of the E-M algorithm.
- **PARAMET** This subroutine does the M-step of the E-M algorithm.
- **ALLOCAT** Allocates individuals into the latent classes according to the value of their posterior probabilities $h(j \mid \mathbf{v}_h, \mathbf{w}_h)$.
- **EXPECT** Computes the expected frequency of each response pattern (binary items only) under the fitted model.

Appendix C

Questionnaire for the two memory data sets

The following questions have to do with how much you remember about the occasion when you first heard of Margaret Thatcher's announcement that she would resign as Prime Minister / Hillsborough football disaster.

Q.1. Taking your answer from this list, how clear is your recollection of the event? [recollect]

Cannot remember it	1
Vague	2
Fairly clear	3
Clear	4
Completely clear	5

Q.2. Thinking back to when you first heard about her resignation/the disaster, can you remember -just answer yes or no-......

	Yes	No
Where you were	1	0
Who you were with	1	0
How you heard about it	1	0
What you were doing	1	0

Appendix D

British Social Attitudes Survey, 1990

These are some of the questions that have been asked in the sexual attitudes section.

Q1...There is law in Britain against sex discrimination, that is against giving unfair preference to men -or to women- in employment, pay and so on. Do you generally support or oppose the idea of a law for this purpose? [SEXLAW]

Q2...Now I would like you to tell me whether, in your opinion, it is acceptable for a homosexual person to be a teacher at a school? [GAYTEASC]

Q3...Now I would like you to tell me whether, in your opinion, it is acceptable for a homosexual person to be a teacher in a college or a university? [GAYTEAHE]

Q4...Now I would like you to tell me whether, in your opinion, it is acceptable for a homosexual person to hold a responsible position in public life? [GAYPUB]

Q5...Do you think female homosexual couples should be allowed to adopt a baby under the same conditions as other couples? [FGAYADPT]

Q6...Do you think male homosexual couples should be allowed to adopt a baby under the same conditions as other couples? [MGAYADPT]

Q7...If a man and a woman have sexual relations before marriage, what would your general opinion be? [BEFORE MARRIAGE]

Q8...What about a married person having sexual relations with someone other than his or her partner? [EXTRA MARITAL]

Q9...What about sexual relations between two adults of the same sex? [SAME SEX]

Appendix E

British Social Attitudes Survey, 1991

These are some of the questions that have been asked in the environment section.
Q.1-7 How concerned are you about each of these environmental issues?

1...insecticides, fertilisers, chemical sprays	
2...thinning of the ozone	Very concerned 4
3...risks from nuclear power stations	A bit concerned 3
4..the greenhouse effect -a rise in the world's temperature	Not very concerned 2
5...using up the earth's remaining coal, oil and gas	Not at all concerned 1
6...the loss of plant and animal species	
7...the transport and disposal of dangerous chemicals	

Q.8-13 How serious an effect on our environment do you think each of these things has?

8 lead from petrol	
9 industrial waste in the rivers and sea	Very Serious 4
10 waste from nuclear electricity stations	Quite Serious 3
11 industrial fumes in the air	Not Very Serious 2
12 acid rain	Not at all Serious 1
13 certain aerosol chemicals in the atmosphere	

Q.14 As far as nuclear power stations are concerned, which of these statements comes closest to your own feelings?

They create very serious risks for the future	4
They create quite serious risks for the future	3
They create only slight risks for the future	2
They create hardly any risks for the future	1

Appendix F

Generalized latent trait models: general form of second derivatives

The second derivatives of the loglikelihood respect to the unknown parameters required in the Newton-Raphson scheme are computed here.

$$
\begin{aligned}
\frac{\partial^2 L}{\partial \alpha_{i0}^2} &= \frac{\partial}{\partial \alpha_{0i}}\{\sum_{t=1}^{k} h(z_t)[\sum_{h=1}^{n} x_{ih}\frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)\alpha(\phi_i)} - \sum_{h=1}^{n}\frac{g(\mathbf{x}_h \mid z_t)}{f(\mathbf{x}_h)\alpha(\phi_i)} b_i'(\theta_i(z_t))]\} \\
&= \sum_{t=1}^{k} h(z_t)\{\sum_{h=1}^{n} x_{ih} g'(\mathbf{x}_h \mid z_t) f(\mathbf{x}_h)\alpha(\phi_i) - \sum_{h=1}^{n} x_{ih} g(\mathbf{x}_h \mid z_t) f'(\mathbf{x}_h)\alpha(\phi_i) - \\
&\quad [\sum_{t=1}^{k}[g'(\mathbf{x}_h \mid z_t) b_i'(\theta_i(z_t)) + g(\mathbf{x}_h \mid z_t) b_i''(\theta_i(z_t))(\theta_i(z_t))'] f(\mathbf{x}_h)\alpha(\phi_i) - \\
&\quad \sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t) f'(\mathbf{x}_h)\alpha(\phi_i) b_i'(\theta_i(z_t))]\}/\{(f(\mathbf{x}_h))^2(\alpha(\phi_i))^2\} \\
&= \sum_{t=1}^{k} h(z_t)\{\sum_{h=1}^{n} x_{ih} g'(\mathbf{x}_h \mid z_t) f(\mathbf{x}_h)\alpha(\phi_i) - \sum_{h=1}^{n} h(z_t) g(\mathbf{x}_h \mid z_t) x_{ih}\alpha(\phi_i)\sum_{t=1}^{k} g'(\mathbf{x}_h \mid z_t) - \\
&\quad \sum_{h=1}^{n} g'(\mathbf{x}_h \mid z_t) f(\mathbf{x}_h)\alpha(\phi_i) b_i'(\theta_i(z_t)) - \sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t) b_i''(\theta_i(z_t))(\theta_i(z_t))' f(\mathbf{x}_h)\alpha(\phi_i) + \\
&\quad \sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t) b_i'(\theta_i(z_t)) \sum_{t=1}^{k} g'(\mathbf{x}_h \mid z_t) h(z_t)\alpha(\phi_i)\}/\{(f(\mathbf{x}_h))^2(\alpha(\phi_i))^2\} \\
&= \sum_{t=1}^{k} h(z_t)\{\sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t)\sum_{t=1}^{k} h(z_t) g'(\mathbf{x}_h \mid z_t)(b_i'(\theta_i(z_t)) - x_{ih}) + \\
&\quad \sum_{h=1}^{n} g'(\mathbf{x}_h \mid z_t) f(\mathbf{x}_h)(x_{ih} - b_i'(\theta_i(z_t))) -
\end{aligned}
$$

$$
\begin{aligned}
& \sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t) b_i''(\theta_i(z_t)) f(\mathbf{x}_h)\}/\{(f(\mathbf{x}_h))^2(\alpha(\phi_i))^2\} \\
= \; & \sum_{t=1}^{k} h(z_t)\{-\sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t) \sum_{t=1}^{k} h(z_t) g(\mathbf{x}_h \mid z_t)(x_{ih} - b_i'(\theta_i(z_t)))^2/\alpha(\phi_i) + \\
& \sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t) f(\mathbf{x}_h)(x_{ih} - b_i'(\theta_i(z_t)))^2/\alpha(\phi_i) - \\
& \sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t) b_i''(\theta_i(z_t)) f(\mathbf{x}_h)\}/\{(f(\mathbf{x}_h))^2(\alpha(\phi_i))^2\} \\
= \; & -\sum_{t=1}^{k} h(z_t) \sum_{h=1}^{n} g(\mathbf{x}_h \mid z_t) b_i''(\theta_i(z_t))/\{f(\mathbf{x}_h \alpha(\phi_i)\}
\end{aligned} \tag{F.1}
$$

Hence,

$$
\begin{aligned}
\frac{\partial^2 L}{\partial \alpha_{i0}^2} &= -\sum_{t=1}^{k} N_t b_i''(\theta_i(z_t))/\alpha(\phi_i) \\
\frac{\partial^2 L}{\partial \alpha_{i1}^2} &= -\sum_{t=1}^{k} z_t N_t b_i''(\theta_i(z_t))/\alpha(\phi_i) \\
\frac{\partial^2 L}{\partial \alpha_{i0} \partial \alpha_{i1}} &= -\sum_{t=1}^{k} N_t b_i''(\theta_i(z_t))/\alpha(\phi_i)
\end{aligned}
$$

Bibliography

Aitkin, M., D. Anderson, and J. Hinde (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, Series A 144*, 419–461.

Aitkin, M. and D. Rubin (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society, Series B 47*, 67–75.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics 21*, 243–247.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika 52*, 317–332.

Albanese, M. and M. Knott (1994). Bootstrapping latent variable models for binary response. *British Journal of Mathematical and Statistical Psychology 47*, 235–245.

Albanese, M. T. (1990). *Latent Variable Models for Binary Response.* Ph. D. thesis, University of London.

Albanese, M. T. and M. Knott (1992). TWOMISS: a computer program for fitting a one- or two- factor logit-probit latent variable model to binary data when observations may be missing. Technical report, Statistics Department, London School of Economics and Political Science, England & Universidade Federal do Rio Grande do Sul, Brazil. The software will be available over EMAIL. Send inquiries to the authors. (EMAIL to M.KNOTT@LSE.AC.UK).

Andersen, E. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of The Royal Statistical Society, Series B 32*, 283–301.

Andersen, E. (1972). The numerical solution of a set of conditional estimation equations. *Journal of The Royal Statistical Society, Series B 34*, 43–54.

Arminger, G. and U. Küsters (1988). Latent trait models with indicators of mixed measurement level. In R. Langeheine and J. Rost (Eds.), *Latent trait and latent class models.* New York: Plenum press.

Arminger, G. and U. Küsters (1989). Construction principles for latent trait models. In C. Clogg (Ed.), *Sociological Methodology*, Volume 19, pp. 369–393. Washington: American Sociological Association.

Barankin, E. and E. Maitra (1963). Generalisation of the Fisher-Darmois-Koopman- Pitman theorem on sufficient statistics. *Sankhyā, A. 25*, 217–244.

Bartholomew, D. (1993). The statistical approach to social measurement. In D. Krebs and P. Schmidt (Eds.), *New Directions in Attitude Measurement.* Berlin: Walter de Gruyter.

Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society, Series B 42*, 293–321.

Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology 34*, 93–99.

Bartholomew, D. J. (1984a). The foundations of factor analysis. *Biometrika 71*, 221–232.

Bartholomew, D. J. (1984b). Scaling binary data using a factor model. *Journal of the Royal Statistical Society, Series B 46*, 120–123.

Bartholomew, D. J. (1987). *Latent Variable Models and Factor Analysis.* London: Charles Griffin & Co. Ltd.

Bartholomew, D. J. (1988). The sensitivity of latent trait analysis to choice of prior distribution. *British Journal of Mathematical and Statistical Psychology 41*, 101–107.

Bartholomew, D. J. (1995). Spearman and the origin and development of factor analysis. *British Journal of Mathematical and Statistical Psychology 48*, 211–220.

Bartlett, M. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of The Royal Statistical Society, Series B 16*, 296–298.

Basilevsky, A. (1994). *Statistical factor analysis and related methods.* New York: John Wiley & Sons.

Bentler, P. and J. Tanaka (1983). Problems with EM algorithms for ML factor analysis. *Psychometrika 48*, 247–251.

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

Bock, R. D. and M. Lieberman (1970). Fitting a response model for n dichotomously scored items. *Psychometrika 35*, 179–197.

Bozdogan, H. (1987). Model selection and Akaike's information criteria (aic). *Psychometrika 52*, 345–370.

Brook, L., B. Taylor, and G. Prior (1991). British social attitudes, 1990, survey. Technical report, SCPR.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Journal of Statistics 8*, 1–24.

Browne, M. W. (1982). Covariance structures. In D. M. Hwakins (Ed.), *Topics in Applied Multivariate Analysis.* Cambridge: University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology 37*, 62–83.

Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika 40*, 5–32.

Clogg, C. C. (1993). Latent class models: recent developments and prospects for the future. In G. Arminger, C. Clogg, and M. Sobel (Eds.), *Handbook of statistical modelling in the social sciences.* New York: Plenum.

Collett, D. (1991). *Modelling binary data.* London: Chapman & Hall.

Collins, L. M., P. Fidler, S. Wugalter, and J. Long (1993). Goodness-of-fit testing for latent class models. *Multivariate Behavioral Research 28*, 375–389.

Cudeck, R. and M. Browne (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research 18*, 147–167.

de Menezes, L. and D. Bartholomew (1996). New developments in latent structure analysis applied to social attitudes. *Journal of the Royal Statistical Society, Series A 159*, 213–224.

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

Everitt, B. (1981). A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distributions. *Multivariate Behavioral Research 16*, 171–180.

Everitt, B. (1988a). A finite mixture model for the clustering of mixed-mode data. *Statistics and Probability Letters 6*, 305–309.

Everitt, B. (1988b). A Monte Carlo investigation of the likelihood ratio test for number of latent classes in latent class analysis. *Multivariate Behavioral Research 23*, 531–538.

Everitt, B. and D. Hand (1981). *Finite mixture distributions.* Great Britain: Chapman and Hall.

Everitt, B. and C. Merette (1990). The clustering of mixed-mode data: a comparison of possible approaches. *Journal of Applied Statistics 17*, 283–297.

Francis, B., M. Green, and C. Payne (Eds.) (1993). *GLIM 4 The statistical system for generalized linear interactive modelling.* Oxford: Oxford University Press.

Gower, J. (1971). A general coefficient of similarity and some of its properties. *Biometrics 27*, 857–872.

Haberman, S. (1977). Maximum likelihood estimates in exponential response models. *Annals of statistics 5*, 815–841.

Holt, J. and G. Macready (1989). A simulation study of the difference chi-square statistic for comparing latent class models under violation of regularity conditions. *Applied Psychological Measurement 13*, 221–231.

Jöreskog, K. G. (1990). New developments in LISREL: analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and*

Quantity 24, 387–404.

Jöreskog, K. G. and D. Sörbom (1988). *PRELIS - A program for multivariate data screening and data summarization. A preprocessor for LISREL* (second ed.). Chicago: Scientific Software, Inc.

Jöreskog, K. G. and D. Sörbom (1993a). *LISREL 8 - User's reference guide.* Chicago: Scientific Software, Inc.

Jöreskog, K. G. and D. Sörbom (1993b). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hove and London: Lawrence Erlbaum associates.

Kashyap, R. (1982). Optimal choice of AR and MA parts in autoregressive moving average models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 4*, 99–104.

Knott, M. and M. T. Albanese (1993). Conditional distributions of a latent variable and scoring for binary data. *Revista Brasileira de Probabilidade e Estatística 6*, 171–188.

Knott, M., M. T. Albanese, and J. Galbraith (1990). Scoring attitudes to abortion. *The Statistician 40*, 217–223.

Lancaster, H. (1954). Traces and cumulants of quadratic forms in normal variables. *Journal of the Royal Statistical Society, Series B 16*, 247–254.

Lawley, D. N. and A. E. Maxwell (1971). *Factor Analysis as a Statistical Method.* London: Butterworth.

Lazarsfeld, P. F. and N. W. Henry (1968). *Latent structure analysis.* New York: Houghton-Mifflin.

Lee, S.-Y., W.-Y. Poon, and P. Bentler (1992). Structural equation models with continuous and polytomous variables. *Psychometrika 57*, 89–105.

Little, R. and D. B. Rubin (1987). *Statistical analysis with missing data.* New York: John Wiley & Sons.

Little, R. and M. Schluchter (1985). Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika 72*, 497–512.

Lord, F. M. and M. R. Novick (1968). *Statistical Theories of Mental Test Scores.* Reading,Mass.: Addison-Wesley.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models* (2nd ed.). London: Chapman & Hall.

Mellenbergh, G. (1992). Generalized linear item response theory. *Psychological Bulletin 115*, 300–307.

Moustaki, I. E. (1995a). CLASSMIX: A computer program for fitting a latent class model to mixed observed items. Technical report, Statistics Department, London School of Economics and Political Science.

Moustaki, I. E. (1995b). LATENT: A computer program for fitting a one- or two- factor latent variable model to mixed observed items. Technical report, Statistics Department, London School of Economics and Political Science.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika 43*, 551–560.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variables indicators. *Psychometrika 49*, 115–132.

Muthén, B. (1987). *LISCOMP: Computer program.* Chicago: Scientific Software.

Muthén, B. and A. Christoffersson (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika 46*, 407–419.

Nelder, J. and R. Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A 135*, 370–384.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coeficient. *Psychometrika 44*, 443–460.

Olsson, U., F. Drasgow, and N. Dorans (1982). The polyserial correlation coeficient. *Psychometrika 47*, 337–347.

Poon, W.-Y. and S.-Y. Lee (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika 52*, 409–430.

Quinn, B., G. McLachlan, and N. Hjort (1987). A note on the Aitkin-Rubin approach to hypothesis testing in mixture models. *Journal of the Royal Statistical Society, Series B 49*, 311–314.

Read, T. and N. Cressie (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York: Springer-Verlag.

Reiser, M. and M. VandenBerg (1994). Validity of the chi-square test in dichotomous variable factor analysis when expected frequencies are small. *British Journal of Mathematical and Statistical Psychology 47*, 85–107.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica 14*, 465–471.

Rubin, D. B. (1976). Inference and missing data (with discussion). *Biometrika 63*, 581–592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons.

Rubin, D. B. and D. T. Thayer (1982). EM algorithms for ML factor analysis. *Psychometrika 47*, 69–76.

Rubin, D. B. and D. T. Thayer (1983). More on EM for ML factor analysis. *Psychometrika 48*, 253–257.

Schepers, A. and G. Arminger (1992). MECOSA. Technical report. Computer Program.

Sclove, S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika 52*, 333–343.

Shea, B. (1984). FACONE: A computer program for fitting the logit latent variable model by maximum likelihood. Technical report, The London School of Economics and Political Science, Department of Statistics, London.

Spearman, C. (1904). 'general intelligence' objectively determined and measured. *American Journal of Psychology 5*, 201–293.

Stroud, A. and D. Secrest (1966). *Gaussian Quadrature Formulas.* Prentice-Hall.

Titterington, D., A. Smith, and U. Makov (1985). *Statistical analysis of finite mixture distributions*. Great Britain: John Wiley & Sons.

Tzamourani, P. and M. Knott (1995). Sensitivity of a logit-probit latent trait model for binary response to the distribution of the factor. Presented at the 9th European meeting of the psychometric society in Leiden.

Witherspoon, S. and J. Martin (1992). What do we mean by green? *British Social Attitudes the 9th report*, 1–26.

Wolfe, J. (1971). A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal populations. Technical Report STB 72-2, U.S. Naval Personnel and Training Research Laboratory, San Diego.

Wright, D. G., G. Gaskell, and C. O'Muircheartaigh (1994). Flashbulb memory assumptions: Using national surveys to explore cognitive phenomena. Technical Report 11, LSE Methodology Institute.

Wu, C. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics 11*, 95–103.