PSYCHOMETRIKA—VOL. 65, NO. 3, 391–411
SEPTEMBER 2000

GENERALIZED LATENT TRAIT MODELS

IRINI MOUSTAKI AND MARTIN KNOTT

LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

In this paper we discuss a general model framework within 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. We discuss in addition the scoring of individuals on the latent dimensions. The general framework presented allows, not only the analysis of manifest variables all of one type but also the simultaneous analysis of a collection of variables with different distributions. The approach used analyzes the data as they are by making assumptions about the distribution of the manifest variables directly.

Key words: generalized linear models, latent trait model, EM algorithm, scoring methods.

1. Introduction

It will be shown in this paper that the latent trait model can be put in a general framework which allows a common methodology to be used for estimating model parameters and scoring individuals on the factor dimensions.

Survey data contains variables which are measured on a binary scale, or a categorical scale (nominal or ordinal), or a metric scale (discrete or continuous) or combinations of the above. A latent trait model fitted to such data must take into account the scale of each variable.

Within the item response theory framework authors such as Lawley and Maxwell (1971), Bock and Aitkin (1981), Bartholomew and Knott (1999), and many other psychometricians have looked at models for either binary or polytomous or metric variables. Bartholomew and Knott proposed a unified estimation method for binary, categorical (nominal) and metric variables. This approach is now put into a generalized linear model framework which also allows simultaneous analysis of all types of variables.

In the underlying variable approach where the observed items are treated as metric through assumed underlying normal variables, important contributions have been made for both single and mixtures of types of variables. That framework covers a wide range of models which also allow relationships among the latent variables and inclusion of exogenous (explanatory) variables. For example Muthén (1984) and Lee, Poon, and Bentler (1992) proposed estimation methods where at the first stage tetrachoric, polychoric (Olsson, 1979) and polyserial (Olsson, Drasgow, & Dorans, 1982) correlations are estimated by ML and at a second stage the model parameters are estimated by weighted least squares using a weight matrix which is supposed to be a consistent estimate of the asymptotic covariance matrix of the correlations estimated at the first stage. This weight matrix grows very rapidly as the number of manifest variables increases and so it is not feasible for large numbers of items. Jöreskog (1990, 1994) discusses the issue of estimating polyserial and polychoric correlation coefficients together with their asymptotic covariance matrix.

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 levels of measurement of any type and metric latent variables.

Requests for reprints should be sent to Irini Moustaki, Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK. E-mail: i.moustaki@lse.ac.uk

0033-3123/2000-3/1998-0630-A $00.75/0

Our approach is an extension of Bartholomew and Knott (1999) for categorical and metric variables and Moustaki (1996) for mixed (binary and metric) manifest variables and it is fundamentally different from the underlying variable approach as will be shown.

Mellenbergh (1992) discusses the issue of putting item response theory in a generalized linear model framework. As he noticed, a latent variable model can be described by a general linear model (GLIM) in which a monotone function of the expected response to a 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.

The GLIM framework for latent variable models has also been discussed by Green (1996). Green allows for "mixed" types of data to be analyzed. The paper briefly refers to the general framework without giving much detail. Scoring methods have also been looked at and some empirical results were also given. The goodness-of-fit of the models was not discussed at all.

Sammel, Ryan, and Legler (1997) also discuss a latent trait model for mixed outcomes with covariate effects both on the observed outcomes and on the latent variables within the generalized linear models framework. Results in their paper are restricted to binary and normal manifest outcomes with one latent variable. In our paper we look only at a measurement model which shows the effect of a set of latent variables on manifest variables from any distribution in the exponential family. In addition to binary and normal outcomes our results cover polytomous, Poisson and gamma distributed variables.

Generalized linear models (GLIM) were introduced by Melder and Wedderburn (1972) and a systematic discussion of them can be found in McCullagh and Nelder (1989). 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 our results the explanatory variables are latent (unobserved) variables.

We put the latent trait model with mixed manifest variables in a general framework and provide the necessary software which will make the proposed theory easily applicable and available to researchers. A program called LATENT (Moustaki, 1999) has been written in FORTRAN 77 to fit all the models proposed in that paper.

Other software such as MULTILOG (Thissen, 1991) and PARSCALE (Muraki & Block, 1991) are available and they can fit a range of response models for nominal responses (Bock, 1972), ordinal responses (Samejima, 1969; Masters, 1982) and multiple choice items (Thissen & Steinberg, 1984). Those programs deal only with unidimensional categorical observed items. The programs provide marginal maximum likelihood item parameter estimates using an EM algorithm for most of the models they fit.

The paper is organized as follows: section 2 discusses the theoretical framework of generalized linear models; section 3 discusses the estimation method used for the generalized latent trait model; section 4 discusses scoring methods; section 5 presents the results for more than one latent variable; section 6 illustrates the methodology using an example from the Eurobarometer survey of 1992 (section on science and technology), finally section 7 outlines the results of this work.

2. Generalized Linear Models

A generalized linear model consists of three components:

1. The random component in which each of the p random response variables, $(x_1, \ldots, x_p)$ has a distribution from the exponential family, (such as Bernoulli, Poisson, Multinomial, Normal, Gamma).
2. The systematic component in which covariates, here the latent variables, $\mathbf{z}' = z_1, z_2, \ldots, z_q$ produce a linear predictor η_i corresponding to each x_i:

$$\eta_i = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j, \quad i = 1, \ldots, p. \tag{1}$$

3. The links between the systematic component and the conditional means of the random component distributions:

$$\eta_i = v_i(\mu_i(\mathbf{z}))$$

where

$$\mu_i(\mathbf{z}) = E(x_i \mid \mathbf{z})$$

and $v_i(\cdot)$ is called the link function which can be any monotonic differentiable function and may be different for different manifest variables $x_i, i = 1, \ldots, p$.

In item response theory where z is usually a scalar and the observed variables binary, the link function used is the logit or the probit and $\mu_i(\mathbf{z})$ is a probability known as response function.

We shall, in fact, assume that $(x_1, x_2, \ldots, x_p)$ denotes a vector of p manifest variables where each variable has a distribution in the exponential family taking the form:

$$f_i(x_i; \theta_i, \phi_i) = \exp\left\{\frac{x_i\theta_i - b_i(\theta_i)}{\phi_i} + d_i(x_i, \phi_i)\right\}, \quad i = 1, \ldots, p, \tag{2}$$

where $b_i(\theta_i)$ and $d_i(x_i, \phi_i)$ are specific functions taking a different form depending on the distribution of the response variable x_i. All the distributions discussed in this paper have canonical link functions with $\theta_i = \eta_i$; ϕ_i is a scale parameter.

We give below for several different types of responses the three components of the generalized model. We will illustrate the models with one latent variable. Theoretically, the results are easily extended to any number of latent variables (see sec. 5). So far computational constraints have allowed the fitting of models with up to two latent variables.

2.1. *Binary Responses*

Let x_i take values 0 and 1. Suppose that the manifest binary variable has a Bernoulli distribution with expected value $\pi_i(z)$. The link function is defined to be the logit, that is,

$$v(\pi_i(z)) = \theta_i(z) = \operatorname{logit} \pi_i(z) = \ln\left(\frac{\pi_i(z)}{1 - \pi_i(z)}\right) = \alpha_{i0} + \alpha_{i1} z,$$

where

$$\pi_i(z) = P(x_i = 1 \mid z) = \frac{\exp^{\theta_i(z)}}{(1 + \exp^{\theta_i(z)})}.$$

Then

$$\begin{aligned} b_i(\theta_i(z)) &= \log(1 + \exp^{\theta_i(z)}), \\ d_i(x_i, \phi_i) &= 0, \\ \phi_i &= 1, \end{aligned}$$

and the conditional probability of x_i is

$$g_i(x_i \mid z) = \pi_i(z)^{x_i}(1 - \pi_i(z))^{1 - x_i}. \tag{3}$$

2.2. Polytomous Items

In the polytomous case the indicator variable x_i is replaced by a vector-valued indicator function with its sth element defined as

$$x_{i(s)} = \begin{cases} 1, & \text{if the response falls in category } s, \text{ for } s = 1, \ldots, c_i \\ 0, & \text{otherwise,} \end{cases}$$

where c_i denotes the number of categories of variable i and $\sum_{s=1}^{c_i} x_{i(s)} = 1$. The response pattern of an individual is written as $\mathbf{x}' = (\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_p)$ of dimension $\sum_i c_i$. In this section we look at only variables measured on a nominal scale.

The single response function of the binary case is now replaced by a set of functions $\pi_{i(s)}(z)$ $(s = 1, \ldots, c_i)$ where $\sum_{s=1}^{c_i} \pi_{i(s)}(z) = 1$.

In the binary case both x_i and θ_i are scalars where in the polytomous case they are vectors. The first category of the polytomous variable is arbitrarily selected to be the reference category. The vector $\boldsymbol{\theta}_i(z)$ is written as

$$\boldsymbol{\theta}'_i(z) = \left\{0, \ln \frac{\pi_{i(2)}(z)}{\pi_{i(1)}(z)}, \ldots, \ln \frac{\pi_{i(c_i)}(z)}{\pi_{i(1)}(z)}\right\}, \quad i = 1, \ldots, p.$$

The canonical parameter $\boldsymbol{\theta}_i(z)$ remains a linear function of the latent variable

$$\boldsymbol{\theta}_i(z) = \boldsymbol{\alpha}_{i0} + \boldsymbol{\alpha}_{i1} z,$$

where $\boldsymbol{\alpha}'_{il} = (\alpha_{il}(1) = 0, \alpha_{il}(2), \ldots, \alpha_{il}(c_i))$, $l = 0, 1$, and

$$\pi_{i(s)}(z) = \frac{\exp^{\alpha_{i0}(s)+\alpha_{i1}(s)z}}{\left(\sum_{r=1}^{c_i} \exp^{\alpha_{i0}(r)+\alpha_{i1}(r)z}\right)},$$

$$b_i(\boldsymbol{\theta}_i(z)) = \log\left(\sum_{r=1}^{c_i} \exp^{\alpha_{i0}(r)+\alpha_{i1}(r)z}\right),$$

$$\phi_i = 1.$$

As $\pi_{i(s)}(z)$ is over-parameterized, we fix the parameters of the first category to zero, $\alpha_{i0(1)} = \alpha_{i1(1)} = 0$.

The conditional distribution of $\mathbf{x}_i$ given z is taken to be the multinomial distribution

$$g_i(\mathbf{x}_i \mid z) = \prod_{s=1}^{c_i} (\pi_{i(s)}(z))^{x_{i(s)}} \tag{4}$$

2.3. Normal Distribution

Let x_i have a normal distribution with marginal mean α_{i0} and variance Ψ_{ii}. The link function of the conditional distribution $x_i \mid z$ is the identity

$$v(\mu_i(z)) = \theta_i(z) = \alpha_{i0} + \alpha_{i1} z,$$

Also,

$$b_i(\theta_i(z)) = \frac{[\theta_i(z)]^2}{2}$$

$$\phi_i = \Psi_{ii},$$

$$g_i(x_i \mid z) = \frac{1}{\sqrt{2\pi\Psi_{ii}}} \exp\left\{-\frac{1}{2\Psi_{ii}}(x_i - \alpha_{i0} - \alpha_{i1} z)^2\right\}. \tag{5}$$

2.4. *Gamma Distribution*

Suppose x_i has a Gamma distribution. The link function is the reciprocal:

$$v(\mu_i(z)) = \theta_i(z) = -\frac{1}{\gamma_i(z)} = \alpha_{i0} + \alpha_{i1} z,$$

and

$$b_i(\theta_i(z)) = -\log(-\theta_i(z)) = -\log\left(\frac{1}{\gamma_i(z)}\right),$$

$$\phi_i = \frac{1}{\nu_i},$$

$$d_i(x_i;\phi_i) = \nu_i \log(\nu_i x_i) - \log x_i - \log \Gamma(\nu_i).$$

Hence,

$$g_i(x_i \mid z) = \frac{\exp\{-\frac{\nu_i}{\gamma_i(z)} x_i\} x_i^{\nu_i - 1}}{\left(\frac{\gamma_i(z)}{\nu_i}\right)^{\nu_i} \Gamma(\nu_i)} \qquad \gamma_i(z) > 0, \quad x_i > 0, \quad \nu_i > 0. \tag{6}$$

The shape parameter for the Gamma distribution is $\nu_i = 1/\phi_i$ and the dispersion parameter is $\gamma_i(z)/\nu_i = \gamma_i(z)\phi_i$. Now, the requirement that the $\gamma_i(z)$ is positive imposes restrictions on the values of the parameter estimates which imply that $\theta_i(z)$ is negative. To overcome this problem in the estimation, the distribution of the latent variable z is taken to be a censored normal distribution, that means, that only a subset of the original range of the variable z is taken into account. In the case of mixed observed variables this solution is not appropriate because the distribution of z will be censored for all types of data.

2.5. *Poisson Distribution*

Let x_i denote a Poisson random variable. The link function is defined through

$$v(\mu_i(z)) = \theta_i(z) = \ln \mu_i(z) = \alpha_{i0} + \alpha_{i1} z,$$

$$b_i(\theta_i(z)) = \exp(\theta_i(z)) = \mu_i(z),$$

$$\phi_i = 1,$$

$$g_i(x_i \mid z) = \frac{\mu_i(z)^{x_i}}{x_i!} \exp^{-\mu_i(z)}, \qquad x_i \geq 0.$$

3. Estimation

We estimate the parameters by maximum likelihood based on 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. For simplicity of exposition we initially take a single latent variable, z.

Under the assumption of local independence the joint distribution of the manifest variables is:

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

$$= \int_{-\infty}^{+\infty} \left[\prod_{i=1}^{p} g_i(x_i \mid z)\right] h(z)\, dz.$$

We do not assume that $g_i(x_i \mid z)$ for all the p items must be of the same type: rather, $g_i(x_i \mid z)$ can be any distribution from the exponential family. The latent variable z is taken to be standard normal and $h(z)$ denotes the standard normal density.

For a random sample of size n the loglikelihood is written as

$$\begin{aligned} L &= \sum_{m=1}^{n} \log f(\mathbf{x}_m) \\ &= \sum_{m=1}^{n} \log \int_{-\infty}^{+\infty} g(\mathbf{x}_m \mid z) h(z)\, dz \\ &= \sum_{m=1}^{n} \log \int_{-\infty}^{+\infty} \left[\prod_{i=1}^{p} \exp\left\{ \frac{x_{im}\theta_i(z)}{\phi_i} - \frac{b_i(\theta_i(z))}{\phi_i} + d_i(\phi_i, x_{im}) \right\} \right] h(z)\, dz. \end{aligned} \tag{7}$$

The unknown parameters are in $\theta_i(z)$ and ϕ_i. We differentiate the log-likelihood given in (7) with respect to the model parameters, α_{i0}, α_{i1} and the scale parameter ϕ_i.

Finding partial derivatives, we have

$$\begin{aligned} \frac{\partial L}{\partial \alpha_{il}} &= \sum_{m=1}^{n} \frac{1}{f(\mathbf{x}_m)} \frac{\partial f(\mathbf{x}_m)}{\partial \alpha_{il}} \\ &= \sum_{m=1}^{n} \frac{1}{f(\mathbf{x}_m)} \int_{-\infty}^{+\infty} g(\mathbf{x}_m \mid z) h(z) \frac{\partial}{\partial \alpha_{il}} \left[\frac{x_{im}\theta_i(z)}{\phi_i} - \frac{b_i(\theta_i(z))}{\phi_i} \right] dz. \end{aligned} \tag{8}$$

The integral in (8) can be approximated by Gauss-Hermite quadrature with weights $h(z_t)$ at abscissae z_t. By interchanging the summation we get

$$\begin{aligned} \frac{\partial L}{\partial \alpha_{il}} &= \sum_{t=1}^{k} h(z_t) \left[\sum_{m=1}^{n} x_{im} \frac{g(\mathbf{x}_m \mid z_t)}{f(\mathbf{x}_m)\phi_i} \frac{\partial \theta_i(z_t)}{\partial \alpha_{il}} - \sum_{m=1}^{n} \frac{g(\mathbf{x}_m \mid z_t)}{f(\mathbf{x}_m)\phi_i} \frac{\partial b_i(\theta_i(z_t))}{\partial \alpha_{il}} \right] \\ &= \sum_{t=1}^{k} [r_{1it}\theta_i'(z_t) - N_t b_i'(\theta_i(z_t))]/\phi_i, \end{aligned} \tag{9}$$

where

$$\begin{aligned} \theta_i'(z_t) &= \frac{\partial \theta_i(z_t)}{\partial \alpha_{il}} \\ b_i'(\theta_i(z_t)) &= \frac{\partial b_i(\theta_i(z_t))}{\partial \alpha_{il}} \\ r_{1it} &= h(z_t) \sum_{m=1}^{n} x_{im} g(\mathbf{x}_m \mid z_t)/f(\mathbf{x}_m) \\ &= \sum_{m=1}^{n} x_{im} h(z_t \mid \mathbf{x}_m), \end{aligned} \tag{10}$$

and

$$\begin{aligned} N_t &= h(z_t) \sum_{m=1}^{n} g(\mathbf{x}_m \mid z_t)/f(\mathbf{x}_m) \\ &= \sum_{m=1}^{n} h(z_t \mid \mathbf{x}_m). \end{aligned} \tag{11}$$

Setting the partial derivatives (see (9)) equal to zero the ϕ_i parameter disappears and we get

$$\frac{\partial L}{\partial \alpha_{il}} = \sum_{t=1}^{k} [r_{1it}\theta_i'(z_t) - N_t b_i'(\theta_i(z_t))] = 0, \tag{12}$$

where the $b_i'(\theta_i(z_t))$ becomes

Binary items:	$b_i'(\theta_i(z_t)) = z_t^l \pi_i(z_t),$	$l = 0, 1;$
Normal metric items:	$b_i'(\theta_i(z_t)) = z_t^l(\alpha_{i0} + \alpha_{i1}z_t),$	$l = 0, 1;$
Gamma metric items:	$b_i'(\theta_i(z_t)) = z_t^l \left(-\frac{1}{\alpha_{i0} + \alpha_{i1}z_t}\right),$	$l = 0, 1;$
Poisson items:	$b_i'(\theta_i(z_t)) = z_t^l \exp(\alpha_{i0} + \alpha_{i1}z_t),$	$l = 0, 1;$

Polytomous items: (12) is written as

$$\frac{\partial L}{\partial \alpha_{il(s)}} = \sum_{t=1}^{k} [r_{1it(s)}\theta_{i(s)}'(z_t) - N_t b_i'(\boldsymbol{\theta}_i(z_t))] = 0,$$

where $b_i'(\boldsymbol{\theta}_i(z_t)) = z_t^l \pi_{i(s)}(z_t), \quad l = 0, 1 \quad s = 2, \ldots, c_i$.

By formulating the model in this way it can be seen that the derivatives of the loglikelihood with respect to the unknown parameters are very easily obtained for any distribution from the exponential family. The only information we need is the first derivative of the specific function b_i.

For Normal continuous items we get explicit formulae for the estimated parameters $\hat{\alpha}_{i0}$ and $\hat{\alpha}_{i1}$. For binary, polytomous, Gamma and Poisson items the ML equations are nonlinear equations. The nonlinear equations can be solved using a Newton-Raphson iterative scheme.

Next we differentiate the loglikelihood with respect to the scale parameter ϕ:

$$\frac{\partial L}{\partial \phi_i} = \sum_{m=1}^{n} \frac{1}{f(\mathbf{x}_m)} \sum_{t=1}^{k} h(z_t) g(\mathbf{x}_m \mid z_t) \left\{ -\frac{x_{im}\theta_i(z_t) - b_i(\theta_i(z_t))}{\phi_i^2} + d_i'(\phi_i, x_{im}) \right\}. \tag{13}$$

By interchanging the summation in (13), setting it equal to zero and solving with respect to ϕ_i^2 we have

$$\hat{\phi}_i^2 = \frac{\sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t]}{\sum_{t=1}^{k} \sum_{m=1}^{n} h(z_t \mid \mathbf{x}_m) d_i'(\hat{\phi}_i, x_{im})}. \tag{14}$$

The function $d_i'(\phi_i, x_{im})$ depends on ϕ_i and so we do not get an explicit form for $\hat{\phi}_i$.

More specifically for the different types of distributions, we have that for the Bernoulli, the multinomial and the Poisson distribution the scale parameter $\phi = 1$. For Normal items the form of $d_i(\phi_i, x_i)$ is given by

$$d_i(\phi_i, x_i) = -\frac{1}{2} \left[\frac{x_i^2}{\phi_i} + \log(2\pi\phi_i) \right] \tag{15}$$

and

$$d_i'(\phi_i, x_i) = \frac{1}{2} \left[\frac{x_i^2}{\phi_i^2} - \frac{1}{\phi_i} \right]. \tag{16}$$

Substituting from (16) into (14) we get

$$\hat{\phi}_i = \hat{\Psi}_{ii} = \frac{1}{\sum_{t=1}^{k} N_t} \sum_{t=1}^{k} [r_{2it} - 2\hat{\alpha}_{i0}r_{1it} - 2\hat{\alpha}_{i1}z_t r_{1it} + (\hat{\alpha}_{i0} + \hat{\alpha}_{i1}z_t)^2 N_t], \tag{17}$$

where

$$r_{1it} = \sum_{m=1}^{n} x_{im} h(z_t \mid \mathbf{x}_m)$$

$$r_{2it} = \sum_{m=1}^{n} x_{im}^2 h(z_t \mid \mathbf{x}_m) \tag{18}$$

$$N_t = \sum_{m=1}^{n} h(z_t \mid \mathbf{x}_m).$$

For Gamma metric variables the form of $d_i(\phi_i, x_i)$ is given by

$$d_i(\phi_i, x_i) = \nu_i \log \nu_i x_i - \log x_i - \log \Gamma(\nu_i) \tag{19}$$

where $\nu_i = 1/\phi_i$. The first derivative of the function $d_i(\phi_i, x_i)$ required by (14) is

$$d_i'(\phi_i, x_i) = -\frac{1}{\phi_i^2} \log \frac{x_i}{\phi_i} - \frac{1}{\phi_i^2} + \frac{1}{\phi_i^2} \frac{1}{\Gamma(\phi_i^{-1})} [\Gamma(\phi_i^{-1})]'. \tag{20}$$

From (14) we get

$$\phi_i^2 = \frac{\sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t]}{\sum_{t=1}^{k} h(z_t) \sum_{m=1}^{n} \frac{g(\mathbf{x}_m|z_t)}{f(\mathbf{x}_m)} \left[-\frac{1}{\phi_i^2} \left[\log \frac{x_{im}}{\phi_i} + 1 - \frac{\Gamma(\phi_i^{-1})'}{\Gamma(\phi_i^{-1})} \right] \right]}. \tag{21}$$

And so

$$\sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t + r_{3it} + N_t] = \log \phi \sum_{t=1}^{k} N_t + \sum_{t=1}^{k} h(z_t) \sum_{m=1}^{n} \frac{g(\mathbf{x}_m \mid z_t)}{f(\mathbf{x}_m)} \frac{\Gamma(\phi_i^{-1})'}{\Gamma(\phi_i^{-1})} \tag{22}$$

where

$$r_{3it} = h(z_t) \sum_{m=1}^{n} (\log x_{im}) g(\mathbf{x}_m \mid z_t)/f(\mathbf{x}_m). \tag{23}$$

But $\Gamma(\phi^{-1})$ does not depend on x_m and so (22) becomes

$$\sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t + r_{3it} + N_t] = \left[\log \phi_i + \frac{\Gamma(\phi_i^{-1})'}{\Gamma(\phi_i^{-1})} \right] \sum_{t=1}^{k} N_t. \tag{24}$$

Multiplying both sides of (24) by (-2):

$$-2 \sum_{t=1}^{k} [r_{1it}\theta_i(z_t) - b_i(\theta_i(z_t))N_t + r_{3it} + N_t] = 2 \left[\log \phi_i^{-1} - \frac{\Gamma(\phi_i^{-1})'}{\Gamma(\phi_i^{-1})} \right] \sum_{t=1}^{k} N_t. \tag{25}$$

The maximum likelihood estimate of ϕ_i is given in (25). The left hand side of (25) is the deviance, $D(\mathbf{x}; \hat{\mu})$, of the latent variable model. A ML estimate for the scale parameter can be obtained by a Newton-Raphson algorithm which requires the calculation of gamma, digamma and trigamma functions. An alternative is to use an approximation to the maximum likelihood estimate that is given in McCullagh and Nelder (1989). Under the condition that ν_i is sufficiently large and so ϕ_i is small, an approximation of the maximum likelihood estimate of the scale parameter is:

$$\hat{\nu}^{-1} = \frac{\bar{D}(6 + \bar{D})}{6 + 2\bar{D}}, \tag{26}$$

where $\bar{D} = D(\mathbf{x}; \hat{\mu})/n$. This approximation is used in the program LATENT for estimating the scale parameter in the gamma case. We found that the other model parameters are not sensitive to changes in the value of the scale parameter.

3.1. *EM Algorithm*

Let the vector of observed variables be $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_L)$ where $\mathbf{x}_j$ is a vector of observed variables all of the same type but the type of variables is different for different values of j, $j = 1, \ldots, L$, where L denotes the number of different types of observed variables to be analyzed. Assume that $\mathbf{x}_1$ denotes binary items, $\mathbf{x}_2$ polytomous items, $\mathbf{x}_3$ normal items, $\mathbf{x}_4$ gamma items and $\mathbf{x}_5$ poisson items.

The maximization of the log-likelihood (7) is done by an EM algorithm. The steps of the algorithm are defined as follows:

Step 1. Choose initial estimates for the model parameters (α_{il} and the scale parameters).
Step 2. E-step: Compute the values $r_{1it}(\mathbf{x}_1)$, $r_{1it(s)}(\mathbf{x}_2)$, $r_{1it}(\mathbf{x}_3)$, $r_{1it}(\mathbf{x}_4)$, $r_{1it}(\mathbf{x}_5)$, $r_{2it}(\mathbf{x}_3^2)$, $r_{3it}(\log \mathbf{x}_4)$ and N_t.
Step 3. M-step: Obtain improved estimates for the parameters by solving the non-linear maximum likelihood equations for Bernoulli, Multinomial, Gamma and Poisson distributed variables and using the explicit equations for Normal distributed variables.
Step 4. Return to Step 2 and continue until convergence is attained.

3.2. *Sampling Properties of the Maximum Likelihood Estimates*

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} and the scale parameters are given by the elements of the inverse of the information matrix at the maximum likelihood solution.

An approximation of the information matrix is given by

$$I(\hat{\boldsymbol{\beta}}) = \left\{ \sum_{m=1}^{n} \frac{1}{f(\mathbf{x}_m)^2} \frac{\partial f(\mathbf{x}_m)}{\partial \boldsymbol{\beta}_j} \frac{\partial f(\mathbf{x}_m)}{\partial \boldsymbol{\beta}_k} \right\}^{-1},$$

where $\boldsymbol{\beta}$ is the vector of the estimated parameters.

Alternatively, the information matrix can be obtained within the EM algorithm, see Louis (1982). In Verhelst and Glas (1993), Louis's method used for the Rasch model. In the program LATENT the first method is used to compute standard errors.

3.3. *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 are well known. For example, given that the number of latent variables has been specified a priori, a likelihood ratio statistic can be used for assessing the goodness-of-fit of the linear factor model. For testing the goodness-of-fit of the latent trait model for binary data and polytomous data either the Pearson X^2 or the likelihood ratio statistic is used. Problems arising from the sparseness of the multi-way contingency tables in the binary case are discussed in Reiser and VandenBerg (1994). Bartholomew and Tzamourani (1999) proposed alternative ways for assessing the goodness-of-fit of the model based on Monte Carlo methods and residual analysis.

For the mixed latent trait model the goodness-of-fit of the one or two-factor latent trait model has been looked at separately for the categorical and the continuous part. For the categorical part of the model, significant information concerning the goodness-of-fit of the model is found in the margins. 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 normal part of the model we check the discrepancies between the sample covariance matrix and the one estimated from the model.

Instead of testing the goodness-of-fit of a specified model we could alternatively use a criterion for selecting among a set of different models. This procedure gives information about the goodness-of-fit for each model in comparison with other models. This is particularly useful for the determination of the number of factors required. In our case we will compare the one factor model with the two factor model. Sclove (1987) gives a review of some of the model selection criteria used in multivariate analysis such as those due to Akaike, Schwarz and Kashap. These criteria take into account the value of the likelihood at the maximum likelihood solution and the number of parameters estimated.

Akaike's criterion for the determination of the order of an autoregressive model in time series has also been used for the determination of the number of factors in factor analysis, see Akaike (1987).

$$\mathrm{AIC} = -2[\max L] + 2m \tag{27}$$

where m is the number of model parameters. The model with the smallest AIC value is taken to be the best one.

4. Scoring Methods for the Generalized Latent Trait Model

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 (though with some risk of bias).

Scoring methods have been proposed in the literature for several latent variable models. 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 $\mathbf{x}$, $(E(z \mid \mathbf{x}))$ can be used to score $\mathbf{x}$.

An alternative method uses the component scores (see Bartholomew 1984). That method avoids the calculation of the posterior mean and the numerical integrations involved. Bartholomew 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 x only through $X = \sum_{i=1}^{p} \alpha_{i1} x_i$; X thus is a Bayesian sufficient statistic for z. The sufficiency of X was noted by Birnbaum (1968) for a fixed effects version of the model. The sufficiency depends on the choice of the response function, it holds for the logit response function 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 (1984) and Knott and Albanese (1993) have shown that for the one factor logit/logit model and the one factor 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.

Here we attempt to give a general framework for the scoring methods for the generalized latent trait model based on the ideas discussed above. We are using the framework used in the paper by Knott and Albanese (1993). Here, we 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

$$\begin{aligned} g(\mathbf{x} \mid z) &= \prod_{i=1}^{p} g_i(x_i \mid z) \\ &= \prod_{i=1}^{p} \exp\left\{ \frac{x_i\theta_i(z) - b_i(\theta_i(z))}{\phi_i} + d_i(x_i, \phi) \right\}, \end{aligned} \tag{28}$$

where $\theta_i(z) = \alpha_{i0} + \alpha_{i1}z$.

Equation (28) becomes

$$g(\mathbf{x} \mid z) = \exp[C_o(\mathbf{x}) + C_1(\mathbf{x})z] \exp\left[-\sum_{i=1}^{p} \frac{b_i(\theta_i(z))}{\phi_i} + \sum_{i=1}^{p} d_i(x_i, \phi_i) \right], \tag{29}$$

where

$$C_o(\mathbf{x}) = \sum_{i=1}^{p} \frac{\alpha_{i0}}{\phi_i} x_i \text{ and } C_1(\mathbf{x}) = \sum_{i=1}^{p} \frac{\alpha_{i1}}{\phi_i} x_i.$$

The conditional distribution of zero responses to all items given the latent variable is

$$g(\mathbf{0} \mid z) = \prod_{i=1}^{p} \exp\left\{ -\frac{b_i(\theta_i(z))}{\phi_i} + d_i(0, \phi) \right\}. \tag{30}$$

From (29) and (30) we have

$$g(\mathbf{x} \mid z) = \exp[C_o(\mathbf{x}) + C_1(\mathbf{x})z] g(\mathbf{0} \mid z) \prod_{i=1}^{p} \frac{\exp(d_i(x_i, \phi_i))}{\exp(d_i(0, \phi_i))}. \tag{31}$$

The joint probability of the manifest variables ($\mathbf{x}$) may be written as

$$\begin{aligned} f(\mathbf{x}) &= \int_{-\infty}^{+\infty} g(\mathbf{x} \mid z) h(z)\, dz \\ &= (\exp(C_o(\mathbf{x}))) \prod_{i=1}^{p} \frac{\exp(d_i(x_i, \phi_i))}{\exp(d_i(0, \phi_i))} f(\mathbf{0}) M_{z|0}(C_1(\mathbf{x})), \end{aligned} \tag{32}$$

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

$$\begin{aligned} h(z \mid \mathbf{x}) &= \frac{g(\mathbf{x} \mid z) h(z)}{f(\mathbf{x})} \\ &= \frac{\exp(C_1(\mathbf{x})z) g(\mathbf{0} \mid z) h(z)}{f(\mathbf{0}) M_{z|0}(C_1(\mathbf{x}))}. \end{aligned} \tag{33}$$

The component scores for different type of variables are:

for binary items: $C_1(\mathbf{x}) = \sum_i \alpha_{i1} x_i$ since $\phi = 1$,

for polytomous items: $C_1(\mathbf{x}) = \sum_i \alpha_{i1(s)} x_{i(s)}$,

for Normal items: $C_1(\mathbf{x}) = \sum_i \frac{\alpha_{i1}}{\Psi_{ii}} x_i$ since $\phi_i = \Psi_{ii}$,

for Gamma items: $C_1(\mathbf{x}) = \sum_i \frac{\alpha_{i1}}{\phi_i} x_i$,

for Poisson items: $C_1(\mathbf{x}) = \sum_i \alpha_{i1} x_i$.

When more than one type of manifest variable is fitted the part which is changed in (33) is that depending on the manifest variables, that is, $C_1(\mathbf{x})$. For the case of five different type of manifest variables

$$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5),$$

where $\mathbf{x}_j$ is a vector of observed variables all of the same type but different for different values of j, $j = 1, \ldots, 5$, we get

$$C_1(\mathbf{x}) = C_1(\mathbf{x}_1) + C_1(\mathbf{x}_2) + C_1(\mathbf{x}_3) + C_1(\mathbf{x}_4) + C_1(\mathbf{x}_5). \tag{34}$$

Hence, the component score for each response pattern/individual of the model with variables $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$ each of different type is

$$\sum_i \alpha_{i1} x_{1i} + \sum_i \alpha_{i1(s)} x_{2i(s)} + \sum_i \frac{\alpha_{i1}}{\Psi_{ii}} x_{3i} + \sum_i \frac{\alpha_{i1}}{\phi_i} x_{4i} + \sum_i \alpha_{i1} x_{5i}. \tag{35}$$

The moment generating function of the conditional distribution of z given $\mathbf{x}$ is

$$\begin{aligned} M_{z|x}(t) &= \int_{-\infty}^{+\infty} \exp(tz) h(z \mid \mathbf{x}) dz \\ &= \int \exp(tz) \frac{\exp[c_1(\mathbf{x})z] h(z \mid 0)}{M_{z|0}(c_1(\mathbf{x}))} dz \\ &= \frac{M_{z|0}(c_1(\mathbf{x}) + t)}{M_{z|0}(c_1(\mathbf{x}))} \end{aligned} \tag{36}$$

The results of Knott and Albanese (1993) for the latent trait model with binary items are extended here 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

$$E(z \mid \mathbf{x}) = K'_{z|0}(c_1(\mathbf{x})) \tag{37}$$

and

$$Var(z \mid \mathbf{x}) = K''_{z|0}(c_1(\mathbf{x})), \tag{38}$$

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 $(c_1(\mathbf{x}))$ if the variance of the conditional distribution of z given that all responses are zero is strictly greater than zero.

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.

5. More than one Latent Variable

The results presented in the previous sections assumed only one latent variable. These results are easily generalized to more than one latent variable.

Let the responses $(x_1, x_2, \ldots, x_p)$ be of different type formats in the exponential family.

The generalized latent trait model is written as

$$v_i(\mu_i(z)) = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} z_j \quad i = 1, \ldots, p, \tag{39}$$

where $\mathbf{z}$ denotes the latent variables and $v_i(\cdot)$ 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}, α_{ij} and ϕ_i is based on the maximization of the log-likelihood of the joint distribution of the manifest variables which under the assumption of local independence is written:

$$f(\mathbf{x}) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(\mathbf{x} \mid \mathbf{z}) h(\mathbf{z})\, d\mathbf{z}, \tag{40}$$

where the log-likelihood for a random sample of size n is

$$L = \sum_{m=1}^{n} \log f(\mathbf{x}_m).$$

The maximization of the loglikelihood is done via the EM algorithm described in a preceding section.

The approximate maximum likelihood equations for the unknown parameters are

$$\frac{\partial L}{\partial \alpha_{il}} = \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} z_{jt_j}^{l} [r(1, i, t_1, \ldots, t_q) - N(t_1, \ldots, t_q) b_i'(\theta_i(z_{t_1}, \ldots, z_{t_q}))] = 0, \tag{41}$$

where l takes the value 0 when the parameter α_{i0} is estimated and the value 1 when the coefficients of the latent variables are estimated and $j = 1, \ldots, q$.

Now $r(1, i, t_1, \ldots, t_q)$ takes a different form depending on the type of the ith manifest variable. The only term to be calculated is the first derivative of the function $b_i(\theta_i(z_{t_1}, \ldots, z_{t_q}))$. In the program LATENT the M-step of the EM algorithm (solution to the non-linear equations given in 41) is done through a Newton-Raphson iterative algorithm.

The scale parameter is estimated as before. An explicit solution exists for the Normal items and a ML estimate or an approximation of it can be used for the Gamma items.

5.1. *Interpretation of the Parameters*

In the case of the binary and the polytomous variables 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 is usually taken to be monotonic nondecreasing in the latent space.

The parameters α_{i0} and α_{ij} for the different type of variables are not directly comparable. That causes a problem when we come to identify which factors are important by looking at the factor loadings.

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

For Normal items α_{ij} denotes the covariance between the manifest variable x_i and the latent variable z_j. By dividing α_{ij} by the square root of the variance of the continuous variable x_i we obtain the correlation between the variable x_i and z_j, that is,

$$\alpha_{ij}^* = \frac{\alpha_{ij}}{(\sum_{j=1}^{q} \alpha_{ij}^2 + \Psi_{ii})^{1/2}}. \tag{42}$$

For binary items following Takane and de Leeuw (1987) or Bartholomew and Knott (1999), we standardize as

$$\alpha_{ij}^* = \frac{\alpha_{ij}}{(\sum_{j=1}^{q} \alpha_{ij}^2 + 1)^{1/2}}. \tag{43}$$

In the polytomous case items are treated as nominal. Hence, each category of a polytomous item can be seen as a binary item, where an individual either belongs to that category or does not. The standardization proposed for the binary items above can be used to standardize the coefficients of the latent variables for each category of the manifest variable, that is,

$$\alpha_{ij(s)}^* = \frac{\alpha_{ij(s)}}{(\sum_{j=1}^{q} \alpha_{ij(s)}^2 + 1)^{1/2}}. \tag{44}$$

The coefficients given in (42), (43), and (44) respectively can be used for a unified interpretation of the factor loadings. The standardization of the parameters brings the interpretation close to factor analysis.

6. Eurobarometer, 1992: Science and Technology Section

To illustrate the methodology developed a dataset from the 1992 Eurobarometer survey will be looked at. The data used are from the Consumer Protection and Perceptions of Science and Technology section. The country chosen is Great Britain. The questions chosen for the analysis are given below.

1. Science and technology are making our lives healthier, easier and more comfortable.
2. Scientific and technological research do not play an important role in industrial development.
3. New technology does not depend on basic scientific research.
4. Scientific and technological research cannot play an important role in protecting the environment and repairing it.
5. The benefits of science are greater than any harmful effects it may have.
6. Thanks to science and technology, there will be more opportunities for the future generations.
7. The application of science and new technology will make work more interesting.

Half of the sample were asked the questions with the following response categories: strongly disagree, disagree to some extent, agree to some extent and strongly agree. For the other half of the sample there was a middle category as well. The questions used here are the ones without the middle category. Missing values have been excluded from the analysis by listwise deletion. That left us with 392 respondents. Items 3, 4 and 5 have been negatively expressed. Items have been recoded so that a high score on any of these items shows a positive attitude towards science and technology.

After looking at the frequency distribution of the 7 items we decided to dichotomize the first two items (agree vs. disagree), to treat Items 3 and 4 as polytomous and Items 5, 6 and 7 as normally distributed. The dichotomization of the first two items is done for the purpose of evaluating the procedures. The number of quadrature points used in the estimation is 48 for the one-factor model and (16×16) for the two-factors models. First a one-factor model is fitted to the seven items. The parameter estimates are given in Table 1. Standard errors are enclosed in parentheses.

The goodness-of-fit of the model is checked by looking at the two- and three-way margins of the binary and the polytomous observed variables. For each cell a chi-square value is calculated ($(O - E)^2/E$). The fit of the one factor model on the margins of the binary variables is very satisfactory but not on the polytomous variables margins, see Table 2.

Figure 1 gives the response functions for each category of Items 3 and 4. The plots show an unclear and unexpected pattern. In the case where the response categories of an item are ordered we would expect that individuals with high values on the z-scale have high probabilities of being in the higher category of that item. A reason for the bad fit could be using only one factor.

Hence, one factor cannot account for the interrelationships among the seven items. The two factor model improved the fit on the margins considerably, see Table 3. The AIC criterion values is also smaller for the two factor model than the one factor model which indicates that the two factor model fits the data better. The parameter estimates for the two factor model are given in Table 4.

Figures 2 and 3 give the response functions for each category of the polytomous items. The pattern now is much clearer. The response probabilities for each of the response categories are more distinct than before.

The aim of the analysis is to identify factors that can explain the interrelationships among the seven observed items. The factor solution is not unique since orthogonal rotations of the factors give the same value of the likelihood. This lack of identification does not create problems in the convergence of the EM algorithm. It does create a problem when it comes to computing standard errors for the parameter estimates of the multiple factor model. Because of the non-unique solution the inversion of the information matrix is not possible. To overcome this problem in the program, the value of one model parameter is constrained to be equal to the ML value before the inversion of the information matrix.

TABLE 1.
Estimates and Standard Errors for the One-Factor Latent Trait Model

Binary		α_{i0}	α_{i1}	$\pi_i(z = 0)$
Item 1		2.47 (0.22)	0.73 (0.21)	0.92
Item 2		1.77 (0.15)	−0.01 (0.17)	0.85
Polytomous	**Cat**	α_{i0}	α_{i1}	$\pi_i(z = 0)$
Item 3	1	0.00	0.00	
Item 3	2	2.01 (0.36)	−1.05 (0.30)	0.24
Item 3	3	2.56 (0.35)	−1.02 (0.31)	0.41
Item 3	4	2.34 (0.35)	−0.90 (0.31)	0.33
Item 4	1	0.00	0.00	
Item 4	2	1.38 (0.26)	−0.83 (0.25)	0.23
Item 4	3	1.85 (0.25)	−0.89 (0.24)	0.37
Item 4	4	1.73 (0.26)	−0.65 (0.24)	0.33
Normal		α_{i0}	α_{i1}	Ψ_{ii}
Item 5		2.84 (0.04)	0.31 (0.05)	0.54 (0.05)
Item 6		2.99 (0.04)	0.61 (0.07)	0.20 (0.07)
Item 7		2.72 (0.05)	0.39 (0.05)	0.50 (0.05)
AIC	5189.916			

TABLE 2.
Polytomous Items: $(O - E)^2/E$ Values for the Two-Way Margins, One-Factor Model

	Item 4			
Item 3	1	2	3	4
1	2.67	0.36	0.12	1.75
2	7.32	2.98	0.24	10.3
3	2.25	0.38	5.70	5.61
4	2.16	5.68	8.87	34.6

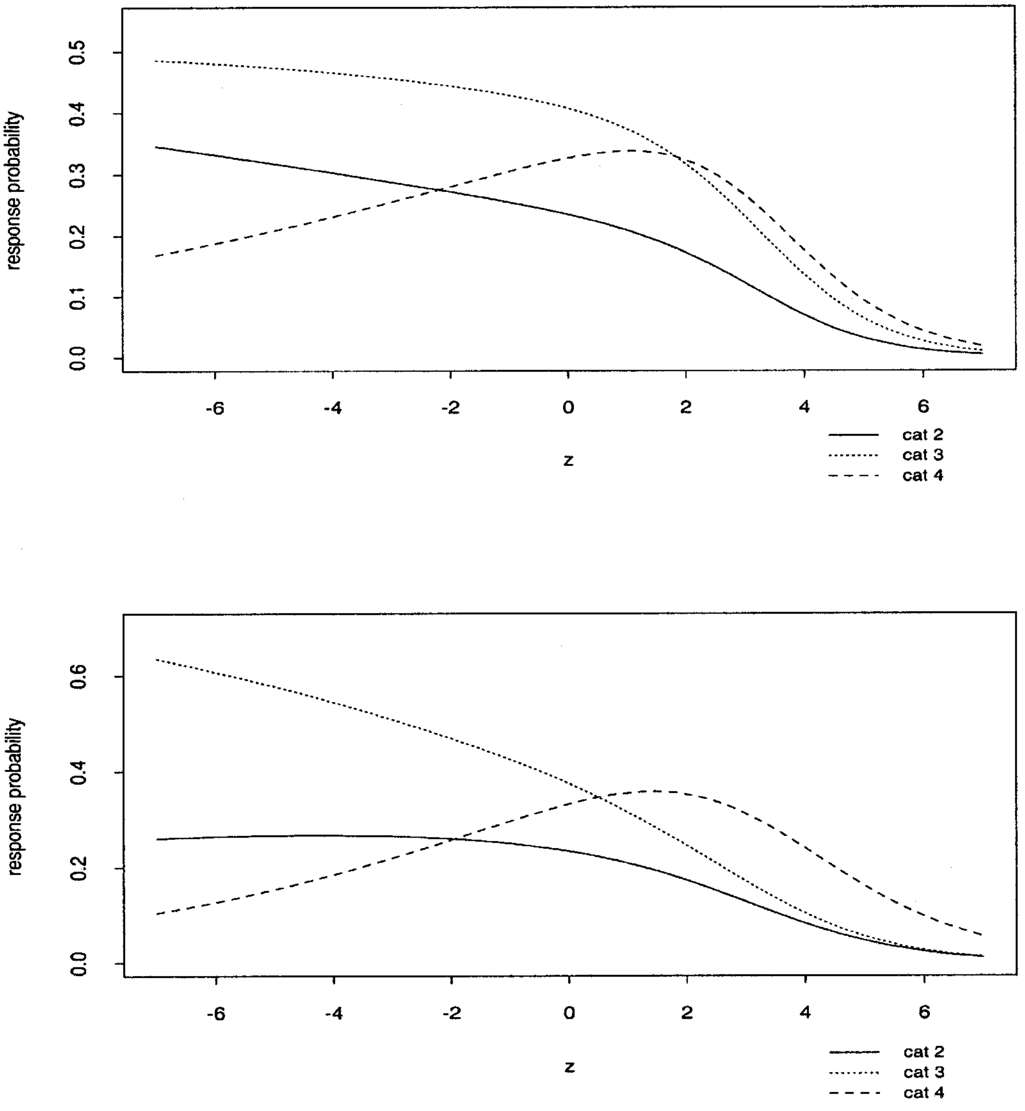

FIGURE 1.
Response functions for item 3 and 4 respectively for the one-factor model.

TABLE 3.
Polytomous Items: $(O - E)^2/E$ Values for the Two-Way Margins, Two-Factor Model

	Item 4			
Item 3	1	2	3	4
1	0.83	0.00	0.53	0.05
2	0.21	0.04	0.23	0.70
3	2.74	0.01	1.76	0.57
4	1.82	0.12	1.07	0.11

With the polytomous items we get parameter estimates (difficulty and discrimination) for each response category. One of the response categories is used as a reference category. The spread of the parameter estimates of the polytomous items indicate that the items have strong discriminating power. The proposed modeling of the polytomous items provide us with information for each category of the item. On the other hand, a parameter for each polytomous item is needed to summarize the discriminating power of the item in relation to the other items in the battery (we will investigate this in further work).

From Table 4 we see that the factor loadings of Items 2, 3 and 4 are larger on the second factor and the opposite is true for Items 1, 5, 6, and 7. Items 2, 3 and 4 are the ones which have been expressed negatively. The question wording might be then one of the reasons that these items form a scale by themselves. An orthogonal rotation of the factor loadings might have suggest a different pattern, but we found the pattern from the unrotated loadings to be the most intuitive one.

Another possible explanation of the need for two dimensions to explain attitudes to scientific and technological research could be that on the one hand people believe that science and technology are important elements of our society but on the other hand people believe that science and technology cannot give an answer to all society's problems.

TABLE 4.
Estimates and Standard Errors for the Two-Factor Latent Trait Model

Binary		α_{i0}	α_{i1}	α_{i2}	$\pi_i(\mathbf{z} = \mathbf{0})$
Item 1		2.63 (0.27)	0.96 (0.27)	0.28 (0.46)	0.93
Item 2		1.98 (0.20)	0.32 (0.39)	0.74 (0.24)	0.88
Polytomous	Cat	$\alpha_{i0(s)}$	$\alpha_{i1(s)}$	$\alpha_{i2(s)}$	$\pi_i(\mathbf{z} = \mathbf{0})$
Item 3	1	0.00	0.00	0.00	
Item 3	2	1.10 (0.91)	−1.55 (0.80)	−0.63 (1.33)	0.12
Item 3	3	3.00 (0.73)	−0.48 (0.95)	1.48 (1.38)	0.83
Item 3	4	−1.45 (12.3)	3.49 (13.3)	10.0 (23.3)	0.01
Item 4	1	0.00	0.00	0.00	
Item 4	2	1.61 (0.30)	−0.58 (0.45)	0.72 (0.39)	0.25
Item 4	3	2.14 (0.28)	−0.55 (0.52)	0.94 (0.37)	0.43
Item 4	4	1.71 (0.34)	0.17(0.93)	2.03 (0.00)	0.28
Normal		α_{i0}	α_{i1}	α_{i2}	Ψ_{ii}
Item 5		2.84 (0.05)	0.29 (0.07)	−0.10 (0.13)	0.54 (0.05)
Item 6		2.99 (0.04)	0.57 (0.12)	−0.25 (0.23)	0.17 (0.07)
Item 7		2.72 (0.05)	0.33 (0.11)	−0.22 (0.14)	0.49 (0.05)
AIC	5104.2				

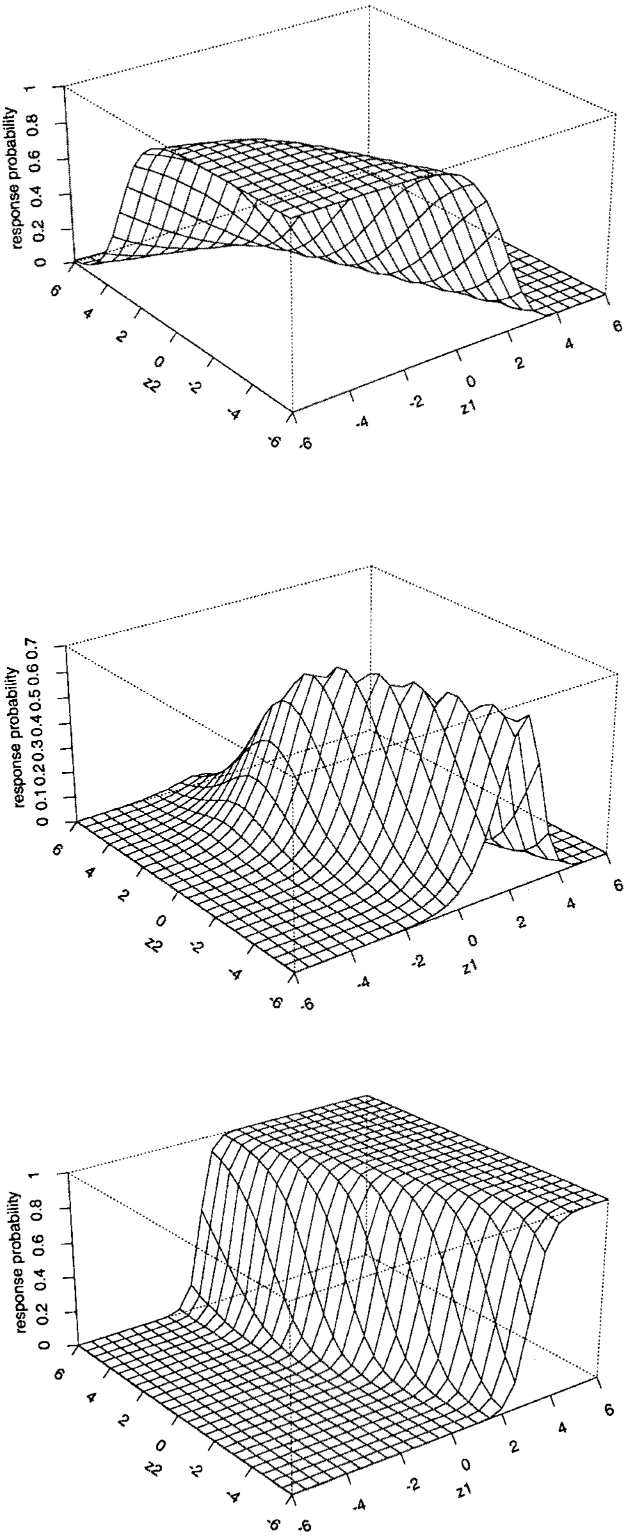

FIGURE 2.
Response function for item 3 and categories 2, 3 and 4 respectively, two-factor model.

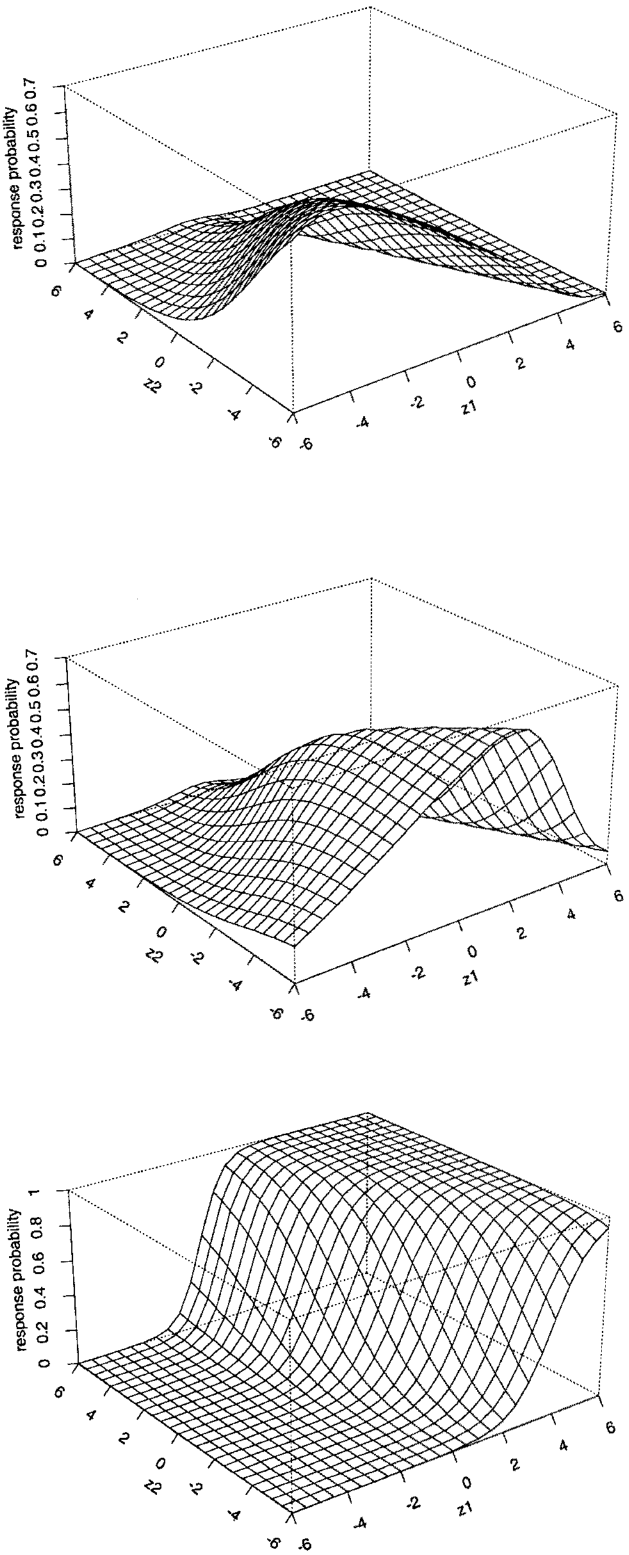

FIGURE 3.
Response function for item 4 and categories 2, 3 and 4 respectively, two-factor model.

7. Conclusion

The methodology developed provides a general framework for both estimating the model parameters of a latent trait model with observed variables of any type from the exponential family and for scoring individuals/response patterns on the identified latent dimensions.

One of the advantages, not discussed in this paper, of having a model which can fit mixed types of variables is that it allows missing values to be incorporated in the analysis. A missing data indicator matrix is used, which contains binary values which indicate when an observation is missing from the data or not. So when the variables to be analyzed are metric or polytomous a model that can handle "mixed" variables is required to fit both the observed part of the data and the missing indicator matrix (see O'Muircheartaigh & Moustaki, 1999). A discussion on how to deal with nonresponse using missing indicator variables can be found in Gifi (1990).

We looked only at polytomous variables on a nominal scale. For the case where the observed variables are on an ordinal scale the canonical link function is not any more a linear function of the parameters. A general class of latent variable models for ordinal variables is discussed in Moustaki (in press).

We have emphasized the interpretation of the model parameters. The standardized coefficients are recommended for a more unified interpretation of the results. The standardization aims to bring the interpretation of the factor coefficients close to factor loadings.

The methods used to examine the goodness-of-fit of the model do not measure the overall fit of the model. However, the chi-square value computed for the one- two- and three-way margins of the binary and polytomous items is an indication of where the model might not fit well. In addition, the AIC is a way of comparing the one factor model with the two factor model. More work needs to be done in establishing an overall goodness-of-fit test for the mixed model.

In this paper we do not discuss the inclusion of covariates effects on the observed variables, however those can be easily incorporated within the generalized linear framework. The explanatory variables can be added as extra terms in the linear predictor equation given in (1) and their effects can be estimated using the EM algorithm presented here.

The software LATENT gives parameter estimates, goodness-of-fit statistics based on the observed and expected frequencies on the one-, two-, and three-way margins of the binary and polytomous items, and scores (component scores, posterior mean) for the individuals in the sample.

The program can handle a large number of observed variables and can fit up to two latent variables. The fit of more than two latent variables is expected to increase the computational time significantly.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317–332.

Arminger, G. & Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 51–73). New York: Plenum Press.

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. (1984). Scaling binary data using a factor model. *Journal of the Royal Statistical Society, Series B, 46*, 120–123.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London: Arnold. (Kendall's Library of Statistics 7)

Bartholomew, D. J., & Tzamourani, P. (1999). The goodness-of-fit of latent trait models in attitude measurement. *Sociological Methods and Research, 27*, 525–546.

Birnbaum, A. (1968). Test scores, sufficient statistics, and the information structures of tests. In F. M. Lord & Novick, M. R. (Eds.), *Statistical theories of mental test scores* (pp. 425–435). Reading, Mass.: Addison-Wesley.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29–51.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of EM algorithm. *Psychometrika, 46*, 443–459.

Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: John Wiley & Sons.
Green, M. (1996). Generalized factor analysis. *Proceedings of the 11th International Workshop on Statistical Modelling*. Orvieto, Italy.
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. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*, 381–389.
Knott, M., & Albanese, M. T. (1993). Conditional distributions of a latent variable and scoring for binary data. *Revista Brasileira de Probabilidade e Estatística*, *6*, 171–188.
Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth.
Lee, S.-Y., Poon, W.-Y., & Bentler, P. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika*, *57*, 89–105.
Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society, Series B*, *44*, 226–233.
Masters, G. N. (1982). A Rasch models for partial credit scoring. *Psychometrika*, *47*, 149–174.
McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.
Mellenbergh, G. (1992). Generalized linear item response theory. *Psychological Bulletin*, *115*, 300–307.
Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology*, *49*, 313–334.
Moustaki, I. (1999). LATENT: A computer program for fitting a one- or two- factor latent variable model to categorical, metric and mixed observed items with missing values (Technical report). London School of Economics and Political Science, Statistics Department.
Moustaki, I. (in press). A latent variable model for ordinal variables. *Applied Psychological Measurement*.
Muraki, E., & Bock, R. D. (1991). *PARSCALE: Parameter scaling of rating data*. Chicago: Scientific Software.
Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variables indicators. *Psychometrika*, *49*, 115–132.
Nelder, J., & Wedderburn, R. (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., Drasgow, F., & Dorans, N. (1982). The polyserial correlation coeficient. *Psychometrika*, *47*, 337–347.
O'Muircheartaigh, C., & Moustaki, I. (1999). Symmetric pattern models: a latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society, Series A*, *162*, 177–194.
Reiser, M., & VandenBerg, M. (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.
Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement No. 17*.
Sammel, M., Ryan, L., & Legler, J. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B*, *59*, 667–678.
Sclove, S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika 52*, 333–343.
Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393–408.
Thissen, D. (1991). *MULTILOG: Multiple, categorical items analysis and test scoring using item response theory*. Chicago: Scientific Software.
Thissen, D., & Steinberg, L. (1984). A model for multiple choice items. *Psychometrika*, *49*, 501–519.
Verhelst, N. D., & Glas, C. A. W. (1993). A dynamic generalization of the rasch model. *Psychometrika*, *58*, 395–415.

Manuscript received 20 JAN 1998
Final version received 13 SEP 1999