Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

British Journal of Mathematical and Statistical Psychology (2008), 61, 415–438
© 2008 The British Psychological Society

The British Psychological Society
www.bpsjournals.co.uk

Generalized latent variable models with non-linear effects

Dimitris Rizopoulos[1] and Irini Moustaki[2,3]*
[1]Erasmus University Medical Center, Rotterdam, The Netherlands
[2]Athens University of Economics and Business, Athens, Greece
[3]London School of Economics and Political Science, UK

Until recently, item response models such as the factor analysis model for metric responses, the two-parameter logistic model for binary responses and the multinomial model for nominal responses considered only the main effects of latent variables without allowing for interaction or polynomial latent variable effects. However, non-linear relationships among the latent variables might be necessary in real applications. Methods for fitting models with non-linear latent terms have been developed mainly under the structural equation modelling approach. In this paper, we consider a latent variable model framework for mixed responses (metric and categorical) that allows inclusion of both non-linear latent and covariate effects. The model parameters are estimated using full maximum likelihood based on a hybrid integration–maximization algorithm. Finally, a method for obtaining factor scores based on multiple imputation is proposed here for the non-linear model.

1. Introduction

The basic idea of latent variable analysis is to find, for a given set of response variables $y_1, \ldots, y_p$, a set of latent variables or factors $z_1, \ldots, z_k$, fewer in number than the observed variables, that contain essentially the same information about dependence. The factors are supposed to account for the dependencies among the response variables in the sense that if the factors were held fixed, the observed variables would be independent.

In the literature, there are two main approaches to the analysis of multivariate data with latent variable models, namely the structural equation modelling (SEM) approach and the item response theory (IRT) approach. There are no differences between SEM and IRT when metric responses are analysed. In the case of categorical responses, the two approaches differ in the way they specify the probability of responding $\Pr(y_1 =$

*Correspondence should be addressed to Dimitris Rizopoulos, Department of Biostatistics, Erasmus University Medical Center, PO Box 2040, CA Rotterdam, The Netherlands (e-mail: d.rizopoulose@erasmusmc.nl).

DOI:10.1348/000711007X213963

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

$a_1, \ldots, y_p = a_p | z_1, \ldots, z_k$) and the way they estimate the model, where $a_1, \ldots, a_p$ represent the different response categories of $y_1, \ldots, y_p$ respectively.

Under the IRT approach, several authors such as Lawley and Maxwell (1971), Bock (1972), Bock and Aitkin (1981) and Bartholomew and Knott (1999) have studied models for binary, polytomous or metric variables. Recently, Bartholomew and Knott (1999), Moustaki and Knott (2000) and Moustaki (2003) proposed a unified full information maximum likelihood (FIML) approach within the generalized linear model framework (McCullagh & Nelder, 1989) that allows simultaneous analysis of all types of variables (nominal, ordinal and metric) leading to a generalized latent variable model (GLVM). The analysis of censored and uncensored variables is discussed in Moustaki and Steele (2005). A similar framework is also discussed by Skrondal and Rabe-Hesketh (2004) which includes multi-level random-effects models as a special case.

Under the SEM approach, ordinal and binary categorical variables are treated as manifestations of underlying normal continuous variables. Different estimation methods have been proposed (see Muthén, 1984; Lee, Poon, & Bentler, 1992; Jöreskog, 1990, 1994; Shi & Lee 1998).

The GLVM that has been considered so far assumes additive latent and covariate effects. However, inclusion of interactions and polynomial terms among latent variables and covariates might be more appropriate in several settings. Applications of latent variable models with non-linear terms in different disciplines such as marketing, social psychology and political theory can be found in Schumacker and Marcoulides (1998) and references therein. The idea of non-linear factor analysis goes back to Gibson (1960) and McDonald (1962, 1967a, 1967b, 1967c). McDonald (1962) discusses cases where manifest variables are not monotonically related to latent variables. The examples given are the typical middle item of an attitude scale and variables that measure the intensity of attitude, which are known to have a U-shape relation to anxiety.

McDonald (1962, 1967a, 1967b, 1967c) has developed an estimation method for fitting factor analysis models with polynomial expressions in the factors (quadratic and interactions between the latent variables). In particular, McDonald (1962) discussed the choice between a linear model with two factors and a model with a single factor and quadratic functions of that single factor. In the linear two-factor model the two factors are assumed to be independent, whereas in the non-linear model the single factor and its quadratic form have correlation zero but are perfectly curvilinearly related. Estimated individual 'factor scores' will give an indication of whether a curvilinear relationship exists between the latent variables or whether two distinct factors are needed. The 'factor scores' estimated are the known principal components scores. His estimation method is based on determining through the relationship between the 'component scores' the parameters defining the best-fit parabola. Estimates of the factor loadings of the non-linear model are obtained either by rotation of the principal component loadings or by a least squares procedure. McDonald (1967a, 1967b, 1967c) extended his non-linear model to allow for interaction terms between the latent variables. Under the strong assumption that the two latent variables are statistically independent, the three components z_1, z_2 and $z_1 \times z_2$ are orthogonal. He considers three different models, a three-factor model where the factors are statistically independent (model 1), a three-factor model with orthogonal factors (model 2) and a model with two factors and the interaction term between them (model 3). The relationships between the factor loadings of models 1 and 2 and the factor loadings of models 2 and 3 are expressed through an orthogonal transformation or through a least squares estimation. The suggested method provides

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

information about whether a weaker model such as model 2 or a stronger model such as model 1 or 3 is required to fit the data. Finally, Etezadi-Amoli and McDonald (1983) proposed a maximum likelihood (ML) ratio and ordinary-least-squares methods for estimating simultaneously factor loadings and factor scores for factor models with both quadratic and interaction terms. In practice, their approach does not allow for a simultaneous solution with respect to factor scores and loadings. They propose a two-stage approach in which, for given values of the latent variables, the estimation of the loadings and any other parameters is obtained by methods applied to polynomial regression models. Etezadi-Amoli and McDonald (1983) have also shown that the non-linear factor model eliminates rotational indeterminacy.

Bartlett (1953) pointed out that the inclusion of the interaction between two latent variables in the linear factor analysis model will produce the same correlation properties of the manifest variables as if a third genuine factor has been included. In addition, Bartlett (1953), Etezadi-Amoli and McDonald (1983) and Bartholomew (1985) suggested graphical ways (bivariate plots of factor scores and components) of investigating the necessity of non-linear terms.

Kenny and Judd (1984) also modelled latent variable interaction terms in one stage under SEM. In their formulation, products of the observed variables are used as indicators for the non-linear terms of the model. Their paper led to many methodological papers discussing several aspects of non-linear latent variable modelling, mainly under the SEM approach. In particular, Jöreskog and Yang (1996) and Yang (1997) studied the limitations of the Kenny–Judd model and found that one product of observed variables is required to identify the model as long as intercept terms and means are included. Problems that arise when using products of observed variables are discussed in Arminger and Muthén (1998). These can be summarized in the difficulty of computing covariance/correlation matrices among non-linear observed variables and by the fact that the distribution of the endogenous latent variables that now includes non-linear terms of exogenous latent variables is complex. This requires the setting of non-linear constraints that complicate the specification of the model. From a practical point of view, it is also true that different combinations of observed variables for constructing the products might lead to different parameter estimates. Bollen (1995) proposed a two-stage least squares method using instrumental variables. Also, Bollen (1996) and Bollen and Paxton (1998) proposed a two-stage least squares method. Wall and Amemiya (2000) proposed a method of moments and Wall and Amemiya (2001) an appended product indicator procedure that produces consistent estimates, while Zhu and Lee (1999) and Arminger and Muthén (1998) discussed a Bayesian estimation method for models with metric data. Klein and Moosbrugger (2000) proposed a method called the latent moderated structural equations approach that looks at conditional distributions of the metric observed variables that follow the normality assumption. Variations of those methods as well as comparisons among them can be found in Schumacker and Marcoulides (1998), Arminger and Muthén (1998), Moulder and Algina (2002), Lee, Song, and Poon (2004) and references therein. Recently, a series of papers by Lee and Zhu (2002), Lee and Song (2004a, 2004b), and Song and Lee (2004, 2006) discuss Bayesian and ML estimation methods for non-linear factor models with mixed data that allow for missing outcomes and hierarchical structures. The heavy integrations involved in the Bayesian framework are overcome by the use of a data augmentation scheme and the use of the Gibbs sampler for sampling from complicated joint distributions. In all those papers, dichotomous and ordinal variables are assumed to be manifestations of underlying normally distributed continuous variables.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

In this paper, we incorporate non-linear terms in the factor model within the GLVM framework presented by Bartholomew and Knott (1999). Bartholomew (1985) shifted the interpretation of the factor analysis model from the unobserved factors to observed statistics known as components that summarize the information conveyed in the manifest variables. Those components for the q-factor model without non-linear terms are q in number and are found to be a weighted sum of the manifest variables. The weights are some function of the factor coefficients. Therefore, q components are a sufficient summary for the manifest variables. In the same paper, Bartholomew stated that in principle there is no problem over incorporating quadratic and interaction terms, but he pointed out that although the inclusion of non-linear terms achieves a more parsimonious model (reduced number of latent dimensions), the number of components is not reduced. Bartholomew and McDonald (1986) pointed out that only in the linear case does the number of latent variables equal the number of components.

The extended GLVM proposed here can easily incorporate non-linear terms without making complex distributional assumptions for the joint distribution of the manifest variables. In this paper, we consider only a measurement model with non-linear terms where latent variables, covariates and their non-linear terms are taken to directly affect the observed response variables. Categorical variables are naturally modelled as members of the exponential family and are not treated as manifestations of underlying normal continuous variables. However, we should note that the FIML method is, as expected, computationally much more intensive than the limited-information methods. This is why we present the models with two latent variables, although the proposed theoretical framework can be extended to handle non-linear terms with more, possibly correlated, latent variables. Model parameters are estimated with the ML method using a hybrid integration–maximization algorithm and standard errors are corrected for model misspecification via the sandwich estimator. Finally, a multiple-imputation-like method is proposed for computing factor scores under the proposed model.

2. Generalized latent variable models with non-linear terms

2.1. Model description

Let $y_1, \ldots, y_p$ denote a set of manifest variables. We assume that the correlations among the manifest variables $\mathbf{y}$ are explained by two latent variables z_1 and z_2, and some observed covariates $\boldsymbol{x}$. The distribution of each y_i, given the two latent variables and the covariates, is assumed to be a member of the exponential family, that is

$$p(y_i|\boldsymbol{z}, \boldsymbol{x}; \theta_i, \phi_i) = \exp\left\{\frac{y_i\theta_i - b_i(\theta_i)}{a_i(\phi_i)} + d_i(y_i, \phi_i)\right\}, \quad i = 1, \ldots, p, \tag{1}$$

where $p(\cdot)$ denotes a probability density function, and θ_i and ϕ_i are the natural and dispersion parameters respectively. The subscript i in the functions $b_i(\cdot)$, $a_i(\cdot)$ and $d_i(\cdot)$ implies that the distribution of each y_i can be a different member of the exponential family, for example binomial, Poisson, normal, gamma, thus allowing for mixed-type response variables.

As has been pointed out by many researchers, linear relationships between the latent variables may not be adequate in many circumstances. Here we consider a more general form of the GLVM which allows polynomial and interaction terms to be added in the linear predictor. In general, we assume that the conditional means of the random

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

component distributions are related to the systematic component generated by the covariates and the latent variables through the formula

$$u_i(\mu_i(\boldsymbol{z})) = \boldsymbol{X}\boldsymbol{\beta}_i^{(x)} + \boldsymbol{Z}\boldsymbol{\beta}_i^{(z)}, \tag{2}$$

where $u_i(\cdot)$ and $\mu_i(\boldsymbol{z})$ denote the link function and the conditional mean $E(y_i|\boldsymbol{z})$ for item $\boldsymbol{i}$ respectively. The matrix $\boldsymbol{X}$ is a design matrix of the observed covariates associated with an item-specific parameter vector $\boldsymbol{\beta}_i^{(x)}$ and $\boldsymbol{Z}$ is a design matrix that is a function of z_1 and z_2, including possibly non-linear terms, with an associated item-specific parameter vector $\boldsymbol{\beta}_i^{(z)}$. Although inclusion of non-linear terms increases the complexity of the linear predictor, their consideration might be necessary to describe more complex relationships between the latent and observed variables.

The need to incorporate such terms in GLVM can arise in several settings. For instance, in educational testing it is reasonable to assume that students' abilities may vary with socio-economic status or race. Justification regarding interactions between latent variables follows in the same sense. Moreover, quadratic latent terms might be useful when a non-linear relationship exists between the factors.

Our approach is advantageous from a numerical point of view for two main reasons. First, the GLVM is based on conditional distributions of observed variables given the latent variables and covariates that keep their form within the known distributions of the exponential family. Second, for the two-factor model with non-linear terms the marginal distribution of y is computed using a double integral instead of a higher-order integration, while allowing for more complex latent structures compared with a two-factor model.

2.2. Parameter estimation and standard errors

The estimation of the model is based on ML, which is known to enjoy good optimality properties for large samples (Cox & Hinkley, 1974). Moreover, in latent variable modelling, as also pointed out in Klein and Moosbrugger (2000), the option to use FIML as opposed to limited-information approaches, typically used for interaction modelling, gives important efficiency and power advantages. In general, for a random sample of size n the log likelihood is written as

$$\ell(\boldsymbol{\alpha}) = \sum_{m=1}^{n} \log p(\boldsymbol{y}_m; \boldsymbol{\alpha}) = \sum_{m=1}^{n} \log \left[\int\int p(\boldsymbol{z}_m) p(\boldsymbol{y}_m | \boldsymbol{z}_m; \boldsymbol{\alpha}) d\boldsymbol{z}_m \right], \tag{3}$$

where $\boldsymbol{\alpha}$ is the vector of all the model parameters (factor loadings and scale parameters). The latent variables $z_m^T = (z_{1m}, z_{2m})$ are assumed to follow a bivariate standard normal distribution, with $\mathrm{Corr}(z_1, z_2) = 0$. For notational convenience, the condition on the covariates is suppressed in the distributional expressions. Under the assumption of conditional independence given the latent variables the manifest variables are taken to be independent,

$$p(\boldsymbol{y}_m | \boldsymbol{z}_m; \boldsymbol{\alpha}) = \prod_{i=1}^{p} p(y_{im} | \boldsymbol{z}_m; \boldsymbol{\alpha}_i). \tag{4}$$

If the EM algorithm (Dempster, Laird, & Rubin, 1977) is to be used for estimating the model parameters, one needs to write down the complete-data log likelihood given by

$$\log p(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\alpha}) = \log p(\boldsymbol{y} | \boldsymbol{z}; \boldsymbol{\alpha}) + \log p(\boldsymbol{z}). \tag{5}$$

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Taking into account (4) and (1), the maximization of the log likelihood is easy since each $p(y_i|\boldsymbol{z})$ is an ordinary generalized linear model (GLM) with linear predictor given by (2).

A direct maximization of (3) can be achieved by solving the score equations $\partial\ell(\boldsymbol{\alpha})/\partial\hat{\boldsymbol{\alpha}} = 0$. In order to derive the score vector under (3), we introduce the following notation: let $\boldsymbol{W} = (\boldsymbol{X}, \boldsymbol{Z})$ denote the full design matrix including both covariates and latent variables, with an associated parameter vector $\boldsymbol{\beta}_i = ([\boldsymbol{\beta}_i^{(x)}]^T, [\boldsymbol{\beta}_i^{(z)}]^T)^T$; thus, w_{ml} denotes the mlth element of $\boldsymbol{W}$ and β_{il} denotes the lth element of $\boldsymbol{\beta}_i$. The superscript T denotes the transpose. The first-order derivatives of (3) with respect to β_{il} take the form

$$\begin{aligned}
\frac{\partial\ell(\boldsymbol{\alpha})}{\partial\beta_{il}} &= \sum_{m=1}^{n'} \frac{\partial}{\partial\beta_{il}} \log\left[\int\!\!\int p(\boldsymbol{z}_m)\cdot p(\boldsymbol{y}_m|z_m;\boldsymbol{\alpha})dz_1dz_2\right] \\
&= \sum_{m=1}^{n'} \int\!\!\int \frac{p(\boldsymbol{z}_m)p(\boldsymbol{y}_m|\boldsymbol{z}_m;\boldsymbol{\alpha})}{p(\boldsymbol{y}_m;\boldsymbol{\alpha})} \frac{\partial}{\partial\beta_{il}}\left[\frac{y_{im}\theta_i(\boldsymbol{z}_m) - b_i(\theta_i(\boldsymbol{z}_m))}{a_i(\phi_i)} + d_i(y_{im},\phi_i)\right]dz_1dz_2 \\
&= \sum_{m=1}^{n'} \int\!\!\int \frac{p(\boldsymbol{z}_m|\boldsymbol{y}_m;\boldsymbol{\alpha})}{a_i(\phi_i)}\left\{y_{im}\frac{\partial\theta_i(\boldsymbol{z}_m)}{\partial\beta_{il}} - \frac{\partial b_i(\theta_i(\boldsymbol{z}_m))}{\partial\beta_{il}}\right\}dz_1dz_2 \\
&= \sum_{m=1}^{n'} \int\!\!\int \frac{[y_{im} - \mu_{im}(\boldsymbol{z}_m)]w_{ml}(\boldsymbol{z}_m)}{\mathrm{Var}(\boldsymbol{y}_i|\boldsymbol{z}_m)u_i'(\mu_{im}(\boldsymbol{z}_m))} p(\boldsymbol{z}_m|\boldsymbol{y}_m;\boldsymbol{\alpha})dz_1dz_2,
\end{aligned} \tag{6}$$

where $u_i'(\cdot)$ denotes the first-order derivative of the link function. Moreover, note that the summation is over the $n' \leq n$ individuals who contribute to the ith item, facilitating the handling of non-balanced data sets.

In connection with (5), it is also useful to observe that (6) represents both the score vector of the observed-data log likelihood (3) and the expected value of the score vector of the complete-data log likelihood (5) with respect to the posterior distribution of the latent variables given the observed variables. This implies that (6) can have a double role. In particular, if (6) is solved with respect to $\boldsymbol{\alpha}$ (i.e. fixing $p(\boldsymbol{z}_m|y_m;\boldsymbol{\alpha})$ to the values of $\boldsymbol{\alpha}$ from the previous iteration), then this corresponds to an EM algorithm whereas if (6) is solved with respect to $\boldsymbol{\alpha}$ without assuming $p(\boldsymbol{z}_m|y_m;\boldsymbol{\alpha})$ fixed, then this corresponds to a maximization of the observed-data log likelihood. This in fact allows an interplay between the EM and the maximization of the observed-data log likelihood for the purpose of evaluating (6).

It is known that the ML estimator of the scale parameter ϕ_i is difficult to compute, except from normal responses, thus leading to the use of alternative methods. For the estimation of the scale parameter in GLVM, see Moustaki and Knott (2000).

Regarding the estimation of standard errors, it has been recognized (Busemeyer & Jones, 1983; Ping, 2005) that the inclusion of non-linear terms could lead to potential underestimation. Moreover, even though ML estimates are asymptotically efficient, latent variable modelling is prone to model misspecification since inference is based on the marginal distribution of the manifest variables and the hypothesized latent structure can only be checked implicitly. Misspecification is even more likely to occur when non-linear terms are present, so a distinction between a third correlated factor and a non-linear term is difficult to be made. Here, we propose a robust estimation for the

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

standard errors of $\hat{\boldsymbol{\alpha}}$ based on the sandwich estimator (see White, 1982). In particular, we use

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\alpha}}) = C(\hat{\boldsymbol{\alpha}}) = H^{-1}(\hat{\boldsymbol{\alpha}})K(\hat{\boldsymbol{\alpha}})H^{-1}(\hat{\boldsymbol{\alpha}}), \tag{7}$$

where $H(\hat{\boldsymbol{\alpha}})$ is the observed information matrix evaluated at the ML estimates, and

$$K(\hat{\boldsymbol{\alpha}}) = D(\hat{\boldsymbol{\alpha}})D(\hat{\boldsymbol{\alpha}})^T,$$

with $D(\hat{\boldsymbol{\alpha}})$ denoting the score vector evaluated at the ML estimates. In particular, for the calculation of $H(\hat{\boldsymbol{\alpha}})$ we have derived an analytical expression for the Hessian matrix under the proposed model. For simplicity, we have considered the case where $\theta_i \equiv \boldsymbol{w}_m^T \boldsymbol{\beta}_i$ and $\boldsymbol{\alpha} \equiv \boldsymbol{\beta}$. Under these conventions, the second-order derivatives of the log likelihood take the form

$$\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_{il} \partial \beta_{kc}} = \frac{\partial}{\partial \beta_{kc}} \sum_{m=1}^{n'} D_{ilm}(\boldsymbol{\beta})$$

$$= \begin{cases} \displaystyle\sum_{m=1}^{n'} \iint \mathcal{B}_m \left\{ \frac{-w_{ml}(\boldsymbol{z}_m) w_{mc}(\boldsymbol{z}_m)}{u_i'(\mu_{im}(\boldsymbol{z}_m))} + \mathcal{A}_{ilm}[\mathcal{A}_{icm} - D_{icm}(\boldsymbol{\beta})] \right\} dz_1 dz_2, & k = i, \\ \displaystyle\sum_{m=1}^{n'} \iint \mathcal{B}_m \{ \mathcal{A}_{ilm}[\mathcal{A}_{kcm} - D_{kcm}(\boldsymbol{\beta})] \} dz_1 dz_2, & k \neq i, \end{cases}$$

where

$$\mathcal{A}_{ilm} = (y_{im} - \mu_{im}(\boldsymbol{z}_m)) w_{ml}(\boldsymbol{z}_m), \quad \mathcal{B}_m = p(\boldsymbol{z}_m | \boldsymbol{y}_m; \boldsymbol{\beta}),$$

and $D_{ilm}(\boldsymbol{\beta})$ denotes the lth component of the score vector for the ith item and mth sample unit. The required integrals are approximated using a 40-point adaptive Gauss–Hermite rule. In the more general case where we work under different link functions and we estimate a scale parameter ϕ_i, numerical derivatives (based on the analytical formula for $D(\boldsymbol{\alpha})$) can be used to approximate the Hessian matrix.

2.3. Constraints and identifiability

A potential risk with the inclusion of non-linear terms, when a small number of items is available, is model overparameterization leading to identifiability problems. The general formulation of the model we have considered so far, including mixed-type responses, hinders the derivation of a general rule for identifiability. Thus, we propose here a conservative approach for ensuring identifiability that would be valid if all the response variables were Gaussian. In particular, for normal responses we can include at most $p(p-1)/2$ factor loadings, accounting for both linear and non-linear terms, since this is the maximum number of parameters that the second-order moments $E[y_i y_j]$, $1 \leq i < j \leq p$, imply. This rule is too restrictive for other types of responses, such as Bernoulli, since higher-order moments are required for the complete specification of the response vector. However, we feel that it is safer to impose this kind of restriction in order to achieve identifiability in all cases.

Furthermore, when insufficient items are available to permit the inclusion of item-specific non-linear terms, cross-item parameter constraints must be imposed which result in one non-linear term being shared by all the items. Such constraints can be easily

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

incorporated under the proposed model since only an extra summation, in the expression for the score vector, is needed, that is,

$$\frac{\partial \ell(\boldsymbol{\alpha})}{\partial \beta_l} = \sum_{m=1}^{n} \int\int p(\boldsymbol{z}_m | \boldsymbol{y}_m; \boldsymbol{\alpha}) \sum_{i \in P_m} \frac{[y_{im} - \mu_{im}(\boldsymbol{z}_m)] w_{ml}(\boldsymbol{z}_m)}{\mathrm{Var}(y_i | \boldsymbol{z}_m) u_i'(\mu_{im}(\boldsymbol{z}_m))} dz_1 dz_2,$$

where β_l now denotes the cross-item non-linear term and P_m is the set of items available for the mth sample unit.

2.4. Extension to correlated latent variables

The model considered so far assumes that two latent variables capture the associations between the manifest variables. Two possible extensions within the GLVM framework are to include more than two latent variables and to assume that these latent variables are correlated. In particular, the model definition presented previously can be extended to assume that $\boldsymbol{Z}$ contains ν latent variables, where $\boldsymbol{z} = (z_1, \ldots, z_\nu)$ follows a multivariate normal distribution with mean zero and correlation matrix $\mathbf{R}$. In this case, it is easier to estimate $\mathbf{R}$ parameterized by ψ, using the EM algorithm. The parameter vector $\boldsymbol{\alpha}$ is now extended to include the parameter ψ of the correlation matrix $\mathbf{R}$ as well. The derivatives with respect to ψ are

$$\frac{\partial \ell(\boldsymbol{\alpha})}{\partial \psi} = \sum_{m=1}^{n} \int \left[\frac{\partial}{\partial \psi} \log p(\boldsymbol{z}_m; \mathbf{R}(\psi)) \right] p(\boldsymbol{z}_m | \boldsymbol{y}_m; \boldsymbol{\alpha}) d\mathbf{z}_m$$

$$= \frac{1}{2} \sum_{m=1}^{n} \mathrm{tr}(-\mathbf{R}^{-1}(\psi)\mathcal{K}) + tr(\mathcal{H}\mathcal{V}_m) + \ddot{\boldsymbol{z}}_m^T \mathcal{H} \ddot{\boldsymbol{z}}_m,$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, $\mathcal{K} = \partial \mathbf{R}(\psi)/\partial \psi$ and $\mathcal{H} = \mathbf{R}^{-1}(\psi)\mathcal{K}\mathbf{R}^{-1}(\psi)$. The elements of the matrix $\mathcal{V}_m$ are the conditional variances of the latent variables $z_{jm}(\mathrm{Var}(z_{jm}|y_m; \boldsymbol{\alpha}))$ and the elements of the vector $\ddot{\boldsymbol{z}}_m$ are the conditional means $(E(z_{jm}|y_m; \boldsymbol{\alpha}))$.

Even though such an extension is mathematically possible, we feel that a model including both non-linear and correlation terms between latent variables might be difficult to estimate and interpret in real data applications, and usually a large number of items and sample units will be required to obtain stable parameter estimates.

3. A hybrid integration–maximization algorithm

Due to the high-dimensional parameter space and the requirement for numerical integration, the estimation of GLVM may be computationally demanding. Observe that even in the simple setting, where all the conditional distributions $p(y_i|\boldsymbol{z})$, $i = 1, \ldots, p$, and $p(\boldsymbol{z})$ follow a normal distribution, the inclusion of non-linear terms does not result in a marginal multivariate normal distribution, and therefore numerical integration cannot be avoided. This has been recognized in the SEM approach by Jöreskog and Yang (1996) and discussed in more detail by Klein and Moosbrugger (2000).

A Gaussian quadrature rule is often used for approximating integrals by a finite weighted sum of the integrand. In order to overcome approximation errors (see Pinheiro & Bates, 1995; Lesaffre & Spiessens, 2001), the adaptive Gauss–Hermite rule is

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

used here. In the adaptive quadrature rule, the integral is approximated according to the expression (for simplicity we have dropped the sample unit subscript)

$$\int\int f(\boldsymbol{z})dz_1dz_2 \approx \det\{\Sigma^{1/2}\}\sum_{k_1=1}^{K_1}\sum_{k_2=1}^{K_2} h_{k_1}h_{k_2}\exp\left(z_k^T z_k\right)f\left(\sqrt{2}\Sigma^{1/2}z_k+\boldsymbol{\mu}\right),$$

where $f(\boldsymbol{z})$ denotes the integrand

$$p(\boldsymbol{z}_m)p(\boldsymbol{y}_m|\boldsymbol{z}_m;\boldsymbol{\alpha})$$

when the log likelihood (3) is computed, and

$$\{p(\boldsymbol{z}_m|\boldsymbol{y}_m;\boldsymbol{\alpha})[y_{im}-\mu_{im}(\boldsymbol{z}_m)]w_{ml}(\boldsymbol{z}_m)\}\{\mathrm{Var}(y_i|\boldsymbol{z}_m)u_i'(\mu_{im}(\boldsymbol{z}_m))\}^{-1}$$

when the score vector (6) is computed, $\boldsymbol{\mu}_z$ is the mode of $f(\boldsymbol{z})$, $\boldsymbol{\Sigma}^{1/2}$ is the Cholesky factor of

$$\boldsymbol{\Sigma} = \left[-\frac{\partial^2 \log f(\boldsymbol{z})}{\partial z \partial \boldsymbol{z}^T}|\boldsymbol{z}=\boldsymbol{\mu}_z\right]^{-1},$$

and h_k and $\boldsymbol{z}_k = (z_{1k_1}, z_{2k_2})$ are the weights and abscissae for the Gauss–Hermite rule. We should also note that an adaptive quadrature rule also works in the case of correlated latent variables, in which case $p(\boldsymbol{z}_m;\mathbf{R})$ should be used instead of $p(\boldsymbol{z}_m)$. Finally, under the above transformation we obtain a function proportional to a $N(0, 2^{-1}\mathbf{I})$ ($\mathbf{I}$ denotes the identity matrix) density, which enhances accuracy since the Gauss–Hermite weight function is proportional to this density.

An important drawback of the quadrature methods is that the number of integrand evaluations required to approximate the integral increases exponentially with the number of latent variables. To overcome this problem, a Monte Carlo integration technique can be used as has been successfully applied in the SEM approach (Lee & Song, 2004a; Lee & Zhu, 2002). Monte Carlo integration is known to lead to good integral approximations with fewer evaluations than the adaptive Gauss–Hermite approach, especially in the case of many latent variables.

A common algorithm used to obtain the ML estimates in GLVMs is the EM algorithm, in which the latent variables $\boldsymbol{z}$ are treated as missing data (Moustaki & Knott, 2000; Lee & Song, 2004a). Although, EM is a stable algorithm leading to a likelihood increase at each iteration, its convergence near the neighbourhood of the maximum can be very slow, requiring many iterations. However, near the maximum, alternatives such as the Newton–Raphson and quasi-Newton Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithms (Lange, 2004) provide faster convergence than the EM. Thus, we propose a hybrid maximization approach, which starts with an EM algorithm for a moderate number of iterations (e.g. 30), and then switches to the quasi-Newton BFGS algorithm, until convergence. The required integrals are approximated by a 20-point adaptive Gauss–Hermite rule. In this approach, the EM is mainly used as a refinement of the starting estimates before beginning the main maximization routine. The BFGS uses approximations to the Hessian based on the score vector and the parameter values, and thus avoids the extra calculations for computing the exact Hessian. Using this hybrid procedure, we expect that a small number of BFGS iterations will be needed, since the EM quickly brings the parameters near to the neighbourhood of the ML estimates. Furthermore, Monte Carlo EM or Markov chain Monte Carlo is better suited to coping with the estimation of the high-dimensional integrals involved as the number of factors increases.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

4. Factor scores

One of the aims of latent variable modelling is to assign scores to the sample members on the latent dimensions identified by the model. In IRT, there are three ways of computing factor scores, namely the ML, Bayes model, also called n maximum *a posteriori* (MAP), and Bayes mean, also-called expected *a posteriori* (EAP). The ML and MAP methods require maximization of $p(\boldsymbol{z}|y;\boldsymbol{\alpha})$ with respect to $\boldsymbol{z}$, whereas the EAP method is the mean of the posterior distribution (Bartholomew, 1981; Bartholomew & Knott, 1999).

A key assumption of all three methods is that the item parameters are fixed quantities known in advance. However, since the true values of the parameters are never available, a common approximation is to use $p(\boldsymbol{z}|\boldsymbol{y};\hat{\boldsymbol{\alpha}})$, with $\hat{\boldsymbol{\alpha}}$ denoting the ML estimates. This approach has two drawbacks: first, in small samples the variability introduced by treating $\hat{\boldsymbol{\alpha}}$ as the vector with the true values is ignored, and second, we should note that we are interested in $p(\boldsymbol{z}|y)$, that is, the posterior distribution integrated over the parameters, and not in $p(\boldsymbol{z}|y;\boldsymbol{\alpha})$. Here, we propose a multiple imputation-like method (Rubin, 1987) for obtaining factor scores and variances that takes those issues into account. In particular, we can write

$$p(\boldsymbol{z}|\boldsymbol{y}) = \int p(\boldsymbol{z},\boldsymbol{\alpha}|\boldsymbol{y})d\boldsymbol{\alpha} \quad = \int p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\alpha})p(\boldsymbol{\alpha}|\boldsymbol{y})d\boldsymbol{\alpha} \tag{8}$$

as the posterior predictive distribution of the latent variables under the complete-data model. Asymptotic Bayesian theory (Cox & Hinkley, 1974) implies that $p(\boldsymbol{\alpha}|y)$ can be approximated by $N(\hat{\boldsymbol{\alpha}}, C(\hat{\boldsymbol{\alpha}}))$, where $C(\hat{\boldsymbol{\alpha}})$ is the covariance matrix of $\hat{\boldsymbol{\alpha}}$ computed from (7). This approximation is well suited to our model, since $p(y|\boldsymbol{z};\boldsymbol{\alpha})$ belongs to the exponential family. To evaluate the integral in (8), we propose the following simulation scheme:

(1) Draw a value, say $\tilde{\boldsymbol{\alpha}}$, from $N(\hat{\boldsymbol{\alpha}}, C(\hat{\boldsymbol{\alpha}}))$.
(2) Maximize the density $p(\mathbf{z}|\mathbf{y};\tilde{\boldsymbol{\alpha}})$ and obtain the mode $\hat{\boldsymbol{z}}(\tilde{\boldsymbol{\alpha}})$ and its associated variance

$$\mathrm{Var}(\hat{\boldsymbol{z}}(\tilde{\boldsymbol{\alpha}})) = \left[-\frac{\partial^2 \log p(\boldsymbol{z}|\boldsymbol{y};\tilde{\boldsymbol{\alpha}})}{\partial \boldsymbol{z}\partial \boldsymbol{z}^T}\Big|_{\mathbf{z}=\hat{\mathbf{z}}(\tilde{\alpha})}\right]^{-1}.$$

(3) Repeat the above steps L times and combine the estimates obtained in each repetition using the formulae of multiple imputation:

$$\hat{\boldsymbol{z}} = \frac{1}{L}\sum_{l=1}^{L}\hat{\boldsymbol{z}}(\tilde{\alpha}^{(l)})$$

$$\mathrm{Var}(\hat{\boldsymbol{z}}) = \frac{1}{L}\sum_{l=1}^{L}\mathrm{Var}(\hat{\boldsymbol{z}}(\tilde{\boldsymbol{\alpha}}^{(l)})) + \frac{1+L^{-1}}{L-1}\sum_{l=1}^{L}[\hat{\boldsymbol{z}}(\tilde{\boldsymbol{\alpha}}^{(l)}-\hat{\boldsymbol{z}})]^T[\hat{\boldsymbol{z}}(\tilde{\boldsymbol{\alpha}}^{(l)}-\hat{\boldsymbol{z}})],$$

where $\tilde{\boldsymbol{\alpha}}^{(l)}$ denotes the lth realization of $\boldsymbol{\alpha}$ from $N(\hat{\boldsymbol{\alpha}}, C(\hat{\boldsymbol{\alpha}}))$.

This procedure explicitly acknowledges the uncertainty about the true values of $\boldsymbol{\alpha}$, by drawing from their approximate distribution, while taking into account the Monte Carlo error. The maximization required in step 2 is easily accomplished using standard

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

numerical techniques, since as the number of items increases, $p(\boldsymbol{z}|\mathbf{y};\tilde{\boldsymbol{\alpha}})$ converges to a concave function with respect to $\boldsymbol{z}$ with a single maximum.

In cases where more information (e.g. specific quantiles, shape) about the posterior distribution is needed, we could easily modify step 2 of the above scheme and instead of maximizing draw a value from this density. In particular, either a Metropolis or a rejection sampling step (e.g. Booth & Hobert, 1999) could be adopted to get a sample of L values from

$$p(\boldsymbol{z}|\boldsymbol{y},\tilde{\boldsymbol{\alpha}}^{(l)}) \propto p(\boldsymbol{y}|\boldsymbol{z};\tilde{\boldsymbol{\alpha}}^{(l)})p(\boldsymbol{z}) \propto \exp\left\{\sum_{i=1}^{p}\left[\frac{y_i\tilde{\theta}_i^{(l)} - b_i(\tilde{\theta}_i^{(l)})}{a_i(\tilde{\phi}_i^{(l)})}\right] - \frac{\boldsymbol{z}^T\boldsymbol{z}}{2}\right\}.$$

These values can then be used to extract any required information about this distribution. The choice of L depends on the quantities we wish to estimate as well as the assumptions about $p(\boldsymbol{z}|y)$ we are willing to make. For instance, if we can assume approximate normality of $p(\boldsymbol{z}|y)$ (which will be the case for large p), then we need only enough repetitions to reliably estimate the first two moments, in which case $L = 20$ or 30 suffices. If we do not wish to restrict to the normality assumption and, moreover, want to estimate extreme quantiles (e.g. calculate 95% considence intervals), then L-values of the order of 500 will be required. Finally, it is straightforward to extend the proposed method for obtaining factor scores to handle correlated latent variables by considering $p(\mathbf{z}_m;\mathbf{R})$ in the place of $p(\mathbf{z}_m)$.

5. Simulation

A simulation study is performed to numerically evaluate the performance of the proposed model. In particular, two non-linear models are considered: One that contains two latent variables and their interaction, and one that contains one latent variable and its quadratic effect. For each non-linear model, data are generated under four scenarios: (i) 10 items and 500 sample units, (ii) 10 items and 1,000 sample units, (iii) 30 items and 500 sample units, and (iv) 30 items and 1,000 sample units. We denote by $\boldsymbol{\beta}_{i0}$ the intercept, with $\boldsymbol{\beta}_{i1}, \boldsymbol{\beta}_{i2}$ the loadings for the first and second factor respectively, and with $\boldsymbol{\beta}_{i3}$ the loadings of the non-linear term ($i = 1, \ldots, p$). The parameter values used are as follows: $\boldsymbol{\beta}_0 = \text{seq}(-2.5, 2.5, p)$, $\boldsymbol{\beta}_1 = \text{seq}(-2, 2, p)$, $\boldsymbol{\beta}_2 = \text{seq}(2, -2, p)$, and $\boldsymbol{\beta}_3 = \text{seq}(-1.5, 1.5, p)$, where p denotes the number of items in the corresponding scenario and seq(a, b, p) denotes a regular sequence from a to b of length p (e.g. $\text{seq}(-2, 2, 5) = -2, -1, 0, 1, 2$). Under each scenario, 1,000 data sets are simulated, assuming dichotomous manifest variables and standard normally distributed latent variables. Tables 1–4 present the bias and root mean square error for each parameter.

We can see that bias is relatively small in all cases and does not seem to be significant in any of them. Bias decreases as sample size and the number of items increase. Moreover, the model with the interaction term shows greater bias than the model with the quadratic effect. This could probably be attributed to the higher complexity of the model with the interaction term.

6. Application

As an illustration of the proposed modelling framework, here we discuss an example taken from a section of the 1990 Workplace Industrial Relations Survey (WIRS) dealing

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Table 1. Simulation results for 10 items and 500 sample units, based on 1,000 data sets under the two-factor model with an interaction term (interaction), and the one-factor model with a quadratic term (quadratic). Each entry *a*/*b*/*c* contains the true parameter value *a*, the bias *b* and the root mean square error *c*

Interaction				Quadratic		
β_{i0}	β_{i1}	β_{i2}	β_{i3}	β_{i0}	β_{i1}	β_{i3}
−2.50/−0.76/4.34	−2.00/−0.43/2.29	2.00/−0.22/1.92	−1.50/−0.10/0.84	−2.50/−0.44/1.56	−2.00/−0.33/1.29	−1.50/−0.23/0.97
−1.94/−0.04/1.92	−1.56/0.04/0.70	1.56/0.09/0.85	−1.17/0.21/1.45	−1.94/−0.07/0.64	−1.56/0.01/0.69	−1.17/−0.01/0.73
− 1.39/0.35/1.75	−1.11/0.60/4.44	1.11/−0.68/3.94	−0.83/−0.37/2.35	−1.39/0.04/0.69	−1.11/0.24/0.98	−0.83/0.38/1.44
−0.83/−0.24/1.90	−0.67/−0.11/0.93	0.67/−0.00/3.45	−0.50/0.03/1.13	−0.83/0.44/1.62	−0.67/−0.33/2.64	−0.50/−0.14/2.16
−0.28/0.10/1.03	−0.22/0.27/1.59	0.22/0.36/1.77	−0.17/0.63/4.21	−0.28/−0.07/1.67	−0.22/−0.11/1.05	−0.17/−0.05/0.61
0.28/0.62/3.92	0.22/0.39/2.20	−0.22/0.18/1.96	0.17/0.11/1.13	0.28/0.04/0.69	0.22/0.10/1.10	0.17/0.07/1.71
0.83/0.04/3.51	0.67/−0.03/1.24	−0.67/−0.05/1.02	0.50/−0.18/1.50	0.83/0.34/2.56	0.67/0.20/2.70	0.50/−0.02/2.43
1.39/−0.34/1.65	1.11/−0.55/4.01	−1.11/−0.22/2.88	0.83/0.08/2.56	1.39/0.05/2.02	1.11/0.02/1.66	0.83/−0.01/1.15
1.94/0.02/1.92	1.56/−0.05/1.10	−1.56/−0.02/2.59	1.17/0.02/1.03	1.94/−0.02/0.91	1.56/0.02/0.98	1.17/0.04/1.22
2.50/0.03/1.04	2.00/0.05/1.64	−2.00/0.05/2.12	1.50/0.04/2.99	2.50/−0.05/1.68	2.00/0.14/2.39	1.50/0.11/2.40

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Table 2. Simulation results for 10 items and 1,000 sample units, based on 1,000 data sets under the two-factor model with an interaction term (interaction), and the one-factor model with a quadratic term (quadratic). Each entry *a*/*b*/*c* contains the true parameter value *a*, the bias *b* and the root mean square error *c*

Interaction				Quadratic		
β_{i0}	β_{i1}	β_{i0}	β_{i1}	β_{i0}	β_{i1}	β_{i0}
−2.50/−0.38/2.05	−2.00/−0.26/1.21	2.00/−0.14/1.00	−1.50/−0.08/0.64	−2.50/−0.35/1.30	−2.00/−0.24/0.96	−1.50/−0.17/0.65
−1.94/−0.03/0.41	−1.56/0.02/0.35	1.56/0.07/0.89	−1.17/0.16/1.27	−1.94/−0.05/0.58	−1.56/0.01/0.51	−1.17/−0.00/0.51
−1.39/0.27/1.32	−1.11/0.40/2.31	1.11/−0.33/2.16	−0.83/−0.19/1.29	−1.39/0.05/0.54	−1.11/0.17/0.79	−0.83/0.23/1.02
−0.83/−0.11/1.06	−0.67/−0.06/0.81	0.67/−0.03/0.80	−0.50/−0.00/0.73	−0.83/0.38/1.31	−0.67/−0.04/2.19	−0.50/0.02/1.65
−0.28/0.06/1.07	−0.22/0.14/1.31	0.22/0.20/1.22	−0.17/0.42/2.30	−0.28/−0.04/1.04	−0.22/−0.05/0.70	−0.17/−0.02/0.54
0.28/0.44/1.85	0.22/0.26/1.26	−0.22/0.13/1.14	0.17/0.05/0.82	0.28/0.03/0.54	0.22/0.04/0.72	0.17/0.04/1.30
0.83/0.01/0.82	0.67/−0.02/0.69	−0.67/−0.05/1.15	0.50/−0.15/1.42	0.83/0.06/1.77	0.67/0.25/2.28	0.50/0.09/2.07
1.39/−0.24/1.49	1.11/−0.38/2.12	−1.11/−0.11/2.09	0.83/0.04/1.36	1.39/0.11/1.61	1.11/0.07/1.06	0.83/0.02/0.82
1.94/−0.01/1.19	1.56/−0.03/0.80	−1.56/−0.02/0.61	1.17/0.03/0.55	1.94/−0.02/0.60	1.56/0.01/0.59	1.17/0.01/0.84
2.50/0.00/1.22	2.00/0.01/1.38	−2.00/0.02/1.35	1.50/0.11/2.13	2.50/−0.05/1.30	2.00/−0.04/1.71	1.50/0.03/2.11

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Table 3. Simulation results for 30 items and 500 sample units, based on 1,000 data sets under the two-factor model with an interaction term (interaction), and the one-factor model with a quadratic term (quadratic). Each entry *a*/*b*/*c* contains the true parameter value *a*, the bias *b* and the root mean square error *c*

Interaction				Quadratic		
β_{i0}	β_{i1}	β_{i0}	β_{i1}	β_{i0}	β_{i1}	β_{i0}
−2.50/−0.29/0.78	−2.00/−0.28/0.76	2.00/−0.22/0.81	−1.50/−0.19/0.66	−2.50/−0.23/1.11	−2.00/−0.23/1.03	−1.50/−0.15/0.89
−2.33/−0.18/0.63	−1.86/−0.14/0.55	1.86/−0.11/0.54	−1.40/−0.10/0.47	−2.33/−0.23/0.82	−1.86/−0.14/0.74	−1.40/−0.10/0.65
−2.16/−0.08/0.43	−1.72/−0.06/0.39	1.72/−0.05/0.36	−1.29/−0.04/0.32	−2.16/−0.10/0.61	−1.72/−0.08/0.57	−1.29/−0.05/0.50
−1.98/−0.02/0.31	−1.59/−0.02/0.29	1.59/−0.00/0.27	−1.19/0.01/0.28	−1.98/−0.03/0.49	−1.59/−0.02/0.46	−1.19/−0.01/0.45
−1.81/0.01/0.29	−1.45/0.02/0.31	1.45/0.03/0.32	−1.09/0.05/0.38	−1.81/0.00/0.43	−1.45/0.01/0.41	−1.09/−0.00/0.40
−1.64/0.06/0.39	−1.31/0.08/0.46	1.31/0.10/0.47	−0.98/0.12/0.54	−1.64/−0.00/0.41	−1.31/−0.01/0.43	−0.98/0.01/0.43
−1.47/0.16/0.55	−1.17/0.15/0.59	1.17/0.20/0.64	−0.88/0.26/0.71	−1.47/0.00/0.46	−1.17/0.03/0.46	−0.88/0.04/0.48
−1.29/0.23/0.73	−1.03/0.33/0.82	1.03/−0.36/0.87	−0.78/−0.31/0.86	−1.29/0.07/0.51	−1.03/0.09/0.52	−0.78/0.08/0.58
−1.12/−0.26/0.85	−0.90/−0.22/0.76	0.90/−0.20/0.76	−0.67/−0.14/0.66	−1.12/0.10/0.65	−0.90/0.13/0.75	−0.67/0.20/0.88
−0.95/−0.14/0.67	−0.76/−0.08/0.62	0.76/−0.12/0.59	−0.57/−0.09/0.53	−0.95/0.17/0.97	−0.76/0.22/1.15	−0.57/0.25/1.15
−0.78/−0.05/0.52	−0.62/−0.03/0.48	0.62/−0.03/0.47	−0.47/−0.02/0.45	−0.78/−0.52/1.99	−0.62/−0.24/1.83	−0.47/−0.13/1.57
−0.60/0.01/0.44	−0.48/0.01/0.44	0.48/0.02/0.45	−0.36/0.02/0.46	−0.60/−0.07/1.35	−0.48/−0.15/1.23	−0.36/−0.12/0.99
−0.43/0.02/0.48	−0.34/0.06/0.53	0.34/0.07/0.56	−0.26/0.08/0.59	−0.43/−0.06/0.83	−0.34/−0.11/0.75	−0.26/−0.06/0.57
−0.26/0.08/0.61	−0.21/0.12/0.66	0.21/0.17/0.69	−0.16/0.18/0.72	−0.26/−0.02/0.52	−0.21/−0.05/0.47	−0.16/−0.01/0.45
−0.09/0.21/0.74	−0.07/0.27/0.82	0.07/0.28/0.85	−0.05/0.32/0.91	−0.09/−0.01/0.44	−0.07/−0.01/0.41	−0.05/−0.00/0.40
0.09/0.36/0.87	0.07/0.31/0.87	−0.07/0.30/0.93	0.05/0.27/0.77	0.09/0.01/0.42	0.07/0.02/0.43	0.05/0.03/0.44
0.26/0.23/0.75	0.21/0.19/0.69	−0.21/0.14/0.68	0.16/0.13/0.63	0.26/0.04/0.45	0.21/0.05/0.50	0.16/0.04/0.51
0.43/0.08/0.58	0.34/0.06/0.53	−0.34/0.06/0.52	0.26/0.03/0.48	0.43/0.05/0.60	0.34/0.05/0.67	0.26/0.07/0.80
0.60/0.03/0.46	0.48/0.01/0.45	−0.48/0.01/0.44	0.36/−0.01/0.44	0.60/0.13/1.03	0.48/0.12/1.24	0.36/0.23/1.50
0.78/−0.02/0.44	0.62/−0.04/0.47	−0.62/−0.05/0.49	0.47/−0.06/0.53	0.78/0.25/1.72	0.62/0.23/1.95	0.47/0.61/2.03
0.95/−0.08/0.55	0.76/−0.12/0.64	−0.76/−0.15/0.61	0.57/−0.14/0.67	0.95/−0.47/1.93	0.76/−0.32/1.79	0.57/−0.15/1.59
1.12/−0.18/0.70	0.90/−0.21/0.73	−0.90/−0.24/0.75	0.67/−0.35/0.82	1.12/0.05/1.35	0.90/−0.07/1.26	0.67/−0.02/1.03
1.29/−0.31/0.82	1.03/−0.38/0.91	−1.03/−0.11/0.99	0.78/−0.02/0.94	1.29/0.00/0.88	1.03/−0.01/0.79	0.78/−0.00/0.65
1.47/−0.06/1.02	1.17/−0.00/0.86	−1.17/−0.04/0.84	0.88/−0.04/0.79	1.47/−0.00/0.59	1.17/0.01/0.54	0.88/−0.01/0.50
1.64/−0.01/0.76	1.31/−0.03/0.70	−1.31/0.03/0.64	0.98/−0.01/0.59	1.64/−0.00/0.46	1.31/−0.01/0.41	0.98/−0.01/0.39
1.81/−0.03/0.57	1.45/−0.03/0.52	−1.45/−0.01/0.47	1.09/−0.00/0.46	1.81/−0.01/0.40	1.45/0.01/0.44	1.09/0.01/0.45
1.98/−0.00/0.45	1.59/0.00/0.44	−1.59/0.01/0.45	1.19/−0.01/0.48	1.98/0.00/0.51	1.59/0.00/0.54	1.19/0.01/0.58
2.16/0.04/0.52	1.72/0.03/0.59	−1.72/0.03/0.61	1.29/−0.00/0.66	2.16/−0.01/0.66	1.72/−0.01/0.73	1.29/0.04/0.88
2.33/0.02/0.70	1.86/0.06/0.75	−1.86/0.03/0.77	1.40/0.01/0.82	2.33/0.08/1.11	1.86/0.05/1.26	1.40/0.08/1.54
2.50/−0.02/0.86	2.00/0.01/0.90	−2.00/0.03/0.94	1.50/0.13/0.98	2.50/0.21/1.73	2.00/0.22/1.87	1.50/0.39/1.98

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Table 4. Simulation results for 30 items and 1,000 sample units, based on 1,000 datasets under the two-factor model with an interaction term (interaction), and the one-factor model with a quadratic term (quadratic). Each entry *a*/*b*/*c* contains the true parameter value *a*, the bias *b* and the root mean square error *c*

Interaction				Quadratic		
β_{i0}	β_{i1}	β_{i0}	β_{i1}	β_{i0}	β_{i1}	β_{i0}
−2.50/−0.18/0.69	−2.00/−0.16/0.58	2.00/−0.15/0.58	−1.50/−0.15/0.54	−2.50/−0.22/0.84	−2.00/−0.15/0.78	−1.50/−0.17/0.68
−2.33/−0.11/0.49	−1.86/−0.09/0.48	1.86/−0.08/0.41	−1.40/−0.06/0.39	−2.33/−0.14/0.63	−1.86/−0.11/0.58	−1.40/−0.12/0.55
−2.16/−0.05/0.34	−1.72/−0.04/0.31	1.72/−0.03/0.29	−1.29/−0.02/0.27	−2.16/−0.08/0.51	−1.72/−0.07/0.49	−1.29/−0.04/0.47
−1.98/−0.02/0.25	−1.59/−0.01/0.24	1.59/0.00/0.23	−1.19/0.01/0.23	−1.98/−0.03/0.44	−1.59/−0.03/0.42	−1.19/−0.02/0.40
−1.81/0.01/0.23	−1.45/0.02/0.25	1.45/0.02/0.26	−1.09/0.03/0.29	−1.81/−0.02/0.39	−1.45/0.00/0.37	−1.09/−0.01/0.36
−1.64/0.03/0.31	−1.31/0.05/0.35	1.31/0.06/0.38	−0.98/0.08/0.43	−1.64/0.01/0.36	−1.31/0.01/0.37	−0.98/0.01/0.38
−1.47/0.10/0.46	−1.17/0.10/0.49	1.17/0.13/0.55	−0.88/0.16/0.58	−1.47/0.03/0.40	−1.17/0.02/0.42	−0.88/0.04/0.44
−1.29/0.19/0.63	−1.03/0.20/0.62	1.03/−0.19/0.79	−0.78/−0.19/0.72	−1.29/0.04/0.47	−1.03/0.08/0.48	−0.78/0.07/0.51
−1.12/−0.17/0.71	−0.90/−0.12/0.65	0.90/−0.11/0.63	−0.67/−0.09/0.64	−1.12/0.10/0.58	−0.90/0.11/0.59	−0.67/0.13/0.64
−0.95/−0.06/0.57	−0.76/−0.07/0.55	0.76/−0.04/0.51	−0.57/−0.01/0.48	−0.95/0.14/0.69	−0.76/0.19/0.78	−0.57/0.20/0.85
−0.78/−0.01/0.45	−0.62/−0.02/0.44	0.62/−0.02/0.41	−0.47/−0.01/0.42	−0.78/−0.22/1.42	−0.62/−0.18/1.27	−0.47/−0.10/1.11
−0.60/−0.01/0.39	−0.48/0.00/0.40	0.48/0.00/0.40	−0.36/0.02/0.41	−0.60/−0.09/0.91	−0.48/−0.07/0.81	−0.36/−0.10/0.70
−0.43/0.04/0.44	−0.34/0.04/0.46	0.34/0.03/0.48	−0.26/0.03/0.50	−0.43/−0.06/0.63	−0.34/−0.05/0.55	−0.26/−0.04/0.51
−0.26/0.06/0.55	−0.21/0.06/0.57	0.21/0.08/0.60	−0.16/0.11/0.63	−0.26/−0.04/0.45	−0.21/−0.03/0.41	−0.16/−0.03/0.39
−0.09/0.12/0.68	−0.07/0.16/0.69	0.07/0.20/0.74	−0.05/0.23/0.75	−0.09/−0.02/0.38	−0.07/−0.01/0.37	−0.05/−0.01/0.36
0.09/0.25/0.79	0.07/0.17/0.71	−0.07/0.19/0.70	0.05/0.16/0.68	0.09/0.02/0.36	0.07/0.02/0.36	0.05/0.02/0.38
0.26/0.12/0.63	0.21/0.10/0.62	−0.21/0.11/0.58	0.16/0.09/0.55	0.26/0.04/0.39	0.21/0.04/0.42	0.16/0.06/0.45
0.43/0.04/0.50	0.34/0.05/0.48	−0.34/0.05/0.46	0.26/0.03/0.44	0.43/0.04/0.48	0.34/0.05/0.52	0.26/0.06/0.61
0.60/0.01/0.42	0.48/0.01/0.41	−0.48/−0.00/0.39	0.36/−0.01/0.40	0.60/0.09/0.73	0.48/0.08/0.81	0.36/0.11/0.97
0.78/−0.01/0.40	0.62/−0.01/0.41	−0.62/−0.03/0.43	0.47/−0.06/0.46	0.78/0.13/1.08	0.62/0.19/1.29	0.47/0.22/1.45
0.95/−0.07/0.48	0.76/−0.07/0.52	−0.76/−0.08/0.56	0.57/−0.13/0.59	0.95/−0.13/1.42	0.76/−0.10/1.28	0.57/−0.01/1.16
1.12/−0.13/0.62	0.90/−0.09/0.63	−0.90/−0.17/0.66	0.67/−0.19/0.71	1.12/0.01/0.96	0.90/0.02/0.89	0.67/0.04/0.75
1.29/−0.22/0.73	1.03/−0.18/0.74	−1.03/−0.04/0.86	0.78/−0.06/0.82	1.29/−0.02/0.67	1.03/0.02/0.61	0.78/0.03/0.56
1.47/−0.07/0.78	1.17/−0.02/0.77	−1.17/−0.06/0.72	0.88/−0.06/0.71	1.47/0.02/0.52	1.17/0.02/0.47	0.88/0.00/0.43
1.64/−0.04/0.64	1.31/−0.02/0.60	−1.31/−0.06/0.56	0.98/−0.03/0.51	1.64/0.02/0.41	1.31/0.00/0.36	0.98/0.00/0.34
1.81/−0.03/0.49	1.45/−0.02/0.44	−1.45/−0.02/0.41	1.09/0.01/0.39	1.81/−0.01/0.35	1.45/0.00/0.37	1.09/−0.01/0.40
1.98/−0.01/0.37	1.59/−0.00/0.38	−1.59/0.01/0.39	1.19/0.02/0.41	1.98/−0.01/0.43	1.59/−0.01/0.48	1.19/−0.02/0.51
2.16/0.03/0.45	1.72/0.02/0.48	−1.72/0.02/0.52	1.29/0.08/0.56	2.16/−0.02/0.56	1.72/−0.03/0.58	1.29/−0.02/0.64
2.33/0.04/0.59	1.86/0.06/0.66	−1.86/0.04/0.68	1.40/0.06/0.70	2.33/−0.03/0.77	1.86/−0.03/0.86	1.40/0.01/1.01
2.50/0.06/0.77	2.00/0.07/0.79	−2.00/0.00/0.84	1.50/0.07/0.83	2.50/0.06/1.08	2.00/0.11/1.29	1.50/0.23/1.48

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

with management–worker consultation in firms, see http://www.niesr.ac.uk/research/WERS98/. A subset of these data is used here that consists of 1,005 firms and concerns non-manual workers. Previous analyses of these data can be found in Bartholomew (1998) and in Bartholomew, Steele, Moustaki, and Galbraith (2008). The questions asked are given below.

> Please consider the most recent change involving the introduction of the new plant, machinery and equipment. Were discussions or consultations of any of the type on this card held either about the introduction of the change or about the way it was to be implemented?
>
> (1) Informal discussion with individual workers.
> (2) Meeting with groups of workers.
> (3) Discussions in established joint consultative committee.
> (4) Discussions in specially constituted committee to consider the change.
> (5) Discussions with the union representatives at the establishment.
> (6) Discussions with paid union officials from outside.

All the six items measure the amount of consultation that takes place in firms at different levels of the firm structure and cover a range of informal to formal types of consultation. Those firms that place a high value on consultation might be expected to use all or most consultation practices. We should mention that the above items were not initially constructed to form a scale and therefore our analysis is completely exploratory.

All items are binary and no covariate information is available. Therefore, we use the GLVM in the case of binary responses that depends only on the latent design matrix **Z**. Assuming that each manifest variable y_i has a Bernoulli distribution with expected value $\pi_i(\boldsymbol{z})$ and using the logit link function: we get

$$\theta_i(\boldsymbol{z}) = \text{logit}(\pi_i(\boldsymbol{z})) = \beta_{i0} + \boldsymbol{Z\beta}_i^z,$$

where

$$\pi_i(\boldsymbol{z}) = \Pr(y_i = 1|\boldsymbol{z}) = \frac{\exp(\theta_i(\boldsymbol{z}))}{1 + \exp(\theta_i(\boldsymbol{z}))}.$$

The previous analysis of the WIRS data revealed that neither the one- nor two-factor model provides a good fit to the data. In particular, the two-factor model improved the fit on the two-way margins but not on the three-way margins (see Bartholomew, 1998; Bartholomew *et al.*, 2008).

The use of two latent variables seems logical for the WIRS data, since the items potentially identify two styles of consultation, namely *formal* and *informal*. We extend the simple two-factor model to allow for an interaction between the two latent variables. However, as stated in Section 2.3, the inclusion of item-specific non-linear terms is not advisable here due to the small number of items. Thus, a model with one cross-item interaction term is fitted. In order to ensure convergence, the model was fitted under a variety of starting values for the majority of which the same solution was obtained with the highest log-likelihood value. The likelihood ratio test statistic between the two-factor and the interaction model equals 93.13, which is highly significant at one degree of freedom. However, the use of a likelihood ratio test for the goodness of fit of the model is not advisable here since many expected frequencies are smaller than 5 and thus the approximation of the distribution of the statistic by a χ^2 distribution may not be appropriate. Thus, the fit in the two- and

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

three-way margins is checked instead. We found that all the two- and three-way residuals have acceptable values, which is less than 1 and 2.18 respectively (according to Bartholomew *et al.*, 2008, values greater than 3 or 4 are indicative of poor fit). The good fit of the model is also apparent from a direct inspection of the fitted frequencies given in Table 6.

The ML estimates, their corresponding standard errors and the Wald-based *p*-values are presented in Table 5. All the loadings of the second factor except that for item 1 are positive and large indicating a general factor relating to amount of consultation which take place in firms. The first factor has a large positive loading for item 1, which is an indicator of informal type of consultation. The estimate for the interaction term has a negative sign indicating that the regression coefficient for the one factor decreases linearly with the values of the other. Figure 1 illustrates more clearly the non-linear relationship between the two latent variables.

Table 5. WIRS data: Parameter estimates, standard errors (in brackets) and Wald-based *p*-values for the interaction model

Items	$\hat{\beta}_{i0}$	*p*-value	$\hat{\beta}_i^{z_1}$	*p*-value	$\hat{\beta}_i^{z_2}$	*p*-value
1	−1.30 (0.29)	<.001	4.17 (0.71)	<.001	−0.15 (0.33)	.645
2	1.13 (0.20)	<.001	−1.92 (0.29)	<.001	2.02 (0.40)	<.001
3	−1.53 (0.15)	<.001	0.81 (0.19)	<.001	1.34 (0.23)	<.001
4	−1.55 (0.11)	<.001	−0.48 (0.21)	.026	0.58 (0.24)	.015
5	−1.12 (0.14)	<.001	0.84 (0.19)	<.001	1.65 (0.25)	<.001
6	−4.01 (0.62)	<.001	2.01 (0.37)	<.001	2.78 (0.60)	<.001

Cross-item interaction term
β_{int}: −1.64 (0.16) *p*-value: <.001

In particular, for items 2–6 we observe that the conditional probability of a positive response increases as the formal consultation increases and the informal consultation decreases, and vice versa. The conditional probability for the first item increases only with the values of the first factor and remains nearly constant with respect to the values of the second. This is probably due to the fact that this item represents the main type of informal consultation and acts as the main identifier of the first factor. In conclusion, the more a firm uses one type of consultation, the smaller the probability of also using the other type.

Finally, Table 6 presents the factor scores for the interaction model calculated using both the ML estimates and multiple imputation scheme described in Section 4 with $L = 20$. We observe that the score estimates are nearly the same for both methods with small discrepancies for some patterns with small observed frequencies. However, the standard error estimates using the multiple imputation method are, in the majority of cases, larger than those based on the ML estimates, since the former takes into account the extra variability for not knowing the true parameter values.

7. Conclusion

In this paper, we extend the GLVM framework for modelling mixed-type responses with latent variables and covariates to allow for interaction terms and polynomial effects.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

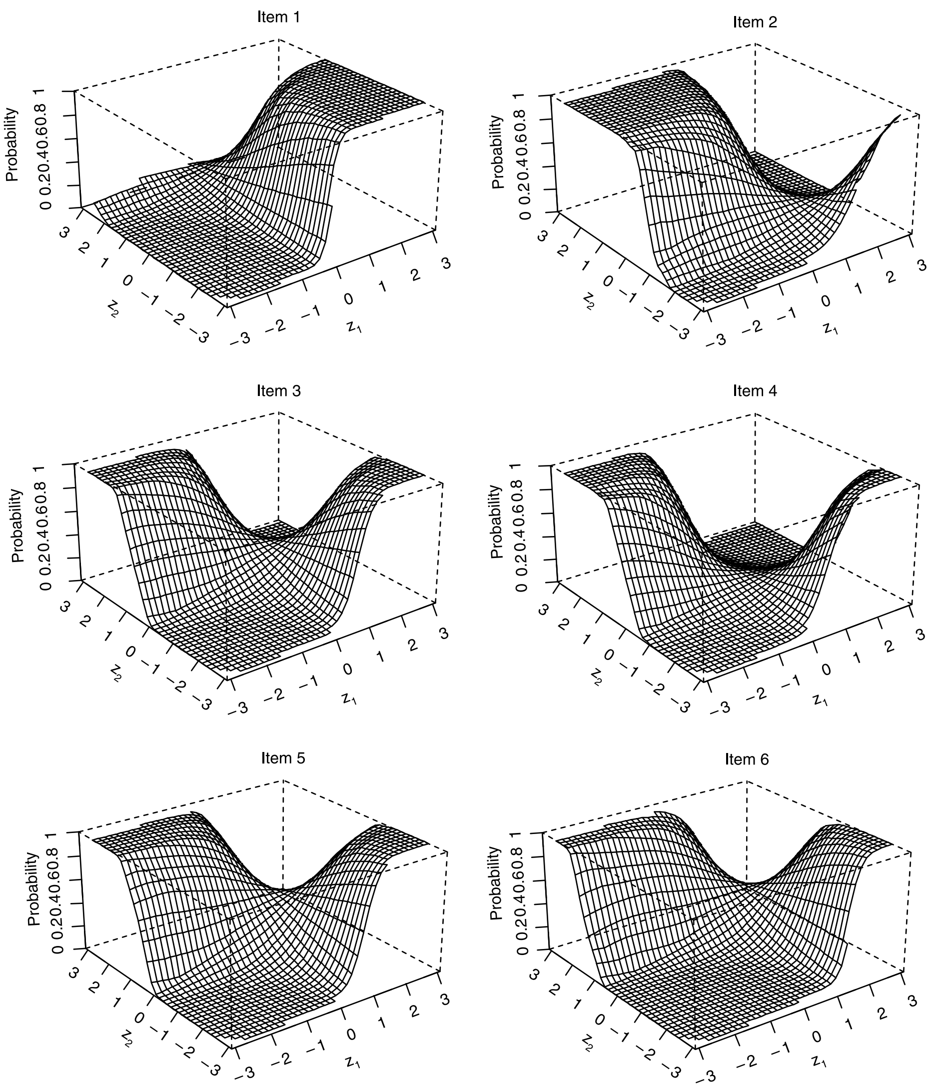

Figure 1. WIRS data: Item characteristic surfaces for items 1–6 based on the interaction model.

The inclusion of non-linear terms allows more complex latent structures to be fitted without substantially increasing the computational burden. In addition, even though it is straightforward to extend the proposed framework to more than two latent variables, we believe that the richness of the two-factor model with non-linear terms will be adequate for the majority of practical applications.

The parameters of the model are estimated with full information maximum likelihood, using a hybrid optimization algorithm. In addition to the usual good properties of maximum likelihood, this estimation procedure also protects against potential bias in the case of ignorable missing-data mechanisms. Moreover, we have considered only adaptive numerical integration to approximate the integrals involved in the calculations of the log likelihood. Schilling and Bock (2005) show that the adaptive

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Table 6. WIRS data: Factor scores based on the ML estimates and multiple imputation (standard errors in brackets) interaction model

Pattern	Observed frequency	Expected frequency	Factor scores using ML estimates		Factor scores using multiple imputation	
			$\hat{z}_1(se(\hat{z}_1))$	$\hat{z}_2(se(\hat{z}_2))$	$\hat{z}_1(se(\hat{z}_1))$	$\hat{z}_2(se(\hat{z}_2))$
000000	132	133.19	0.35 (0.62)	−1.00 (0.61)	0.35 (0.64)	−1.00 (0.63)
100000	65	78.18	−0.45 (0.39)	−0.84 (0.84)	−0.46 (0.40)	−0.82 (0.85)
010000	172	168.73	0.57 (0.71)	−0.22 (0.53)	0.55 (0.71)	−0.22 (0.54)
110000	55	42.33	−0.39 (0.39)	−0.15 (0.71)	−0.40 (0.40)	−0.17 (0.73)
001000	12	7.97	−0.12 (0.40)	−0.32 (0.67)	−0.13 (0.42)	−0.27 (0.71)
101000	24	24.87	−0.88 (0.54)	−0.36 (0.87)	−0.87 (0.54)	−0.32 (0.89)
011000	21	27.15	0.24 (0.59)	0.37 (0.54)	0.25 (0.60)	0.37 (0.54)
111000	13	14.14	−0.53 (0.41)	0.15 (0.72)	−0.53 (0.43)	0.15 (0.74)
000100	13	10.91	0.11 (0.44)	−0.65 (0.61)	0.11 (0.46)	−0.65 (0.63)
100100	20	13.31	−0.58 (0.40)	−0.95 (0.82)	−0.58 (0.41)	−0.94 (0.82)
010100	45	45.97	0.86 (0.82)	0.22 (0.45)	0.84 (0.82)	0.20 (0.46)
110100	5	7.88	−0.36 (0.39)	−0.17 (0.73)	−0.37 (0.42)	−0.20 (0.78)
001100	1	1.32	−0.11 (0.40)	−0.16 (0.64)	−0.11 (0.42)	−0.13 (0.66)
101100	2	5.83	−1.05 (0.42)	−1.10 (0.85)	−1.02 (0.48)	−1.01 (0.89)
011100	13	14.46	0.86 (0.73)	0.66 (0.38)	0.86 (0.74)	0.66 (0.38)
111100	3	3.25	−0.42 (0.51)	0.12 (0.88)	−0.39 (0.65)	0.15 (1.13)
000010	17	12.04	−0.15 (0.42)	−0.17 (0.69)	−0.17 (0.45)	−0.14 (0.75)
100010	34	36.37	−0.93 (0.55)	−0.12 (0.87)	−0.92 (0.56)	−0.12 (0.89)
010010	38	46.18	0.23 (0.59)	0.46 (0.53)	0.22 (0.59)	0.44 (0.54)
110010	26	23.16	−0.52 (0.42)	0.30 (0.70)	−0.53 (0.43)	0.28 (0.73)
001010	2	3.27	−0.38 (0.44)	0.36 (0.71)	−0.38 (0.46)	0.38 (0.76)
101010	31	20.55	−1.31 (0.53)	−0.44 (0.78)	−1.25 (0.61)	−0.32 (0.87)
011010	28	23.78	0.28 (0.58)	0.89 (0.47)	0.28 (0.59)	0.89 (0.48)
111010	6	11.60	−0.50 (0.47)	0.55 (0.76)	−0.52 (0.53)	0.54 (0.85)
000110	0	1.97	−0.11 (0.41)	−0.03 (0.63)	−0.12 (0.43)	−0.03 (0.65)
100110	5	7.03	−1.04 (0.48)	−0.87 (0.88)	−1.00 (0.53)	−0.80 (0.92)
010110	22	25.82	0.84 (0.71)	0.71 (0.38)	0.82 (0.72)	0.70 (0.38)
110110	4	4.85	−0.34 (0.51)	0.34 (0.81)	−0.35 (0.62)	0.32 (0.97)

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Table 6. (*Continued*)

Pattern	Observed frequency	Expected frequency	Factor scores using ML estimates		Factor scores using multiple imputation	
			$\hat{z}_1(se(\hat{z}_1))$	$\hat{z}_2(se(\hat{z}_2))$	$\hat{z}_1(se(\hat{z}_1))$	$\hat{z}_2(se(\hat{z}_2))$
001110	1	0.61	−0.17 (0.46)	0.40 (0.64)	−0.16 (0.48)	0.41 (0.68)
101110	8	7.71	−1.47 (0.37)	−1.30 (0.68)	−1.47 (0.46)	−1.24 (0.71)
011110	35	29.45	1.02 (0.68)	1.06 (0.35)	1.01 (0.68)	1.05 (0.36)
111110	4	4.40	0.22 (0.58)	1.14 (0.50)	0.18 (0.67)	1.10 (0.65)
000001	5	3.29	−1.19 (0.40)	1.72 (0.72)	−1.05 (0.60)	1.41 (1.02)
100001	11	12.49	−1.43 (0.59)	0.76 (0.75)	−1.43 (0.59)	0.71 (0.80)
010001	4	6.02	−0.29 (0.59)	1.02 (0.75)	−0.25 (0.57)	0.94 (0.74)
110001	10	4.96	−0.73 (0.41)	0.85 (0.69)	−0.75 (0.45)	0.75 (0.81)
001001	1	0.89	−0.85 (0.43)	1.27 (0.78)	−0.80 (0.50)	1.15 (0.84)
101001	8	5.59	−1.61 (0.66)	0.08 (0.74)	−1.48 (0.65)	0.22 (0.85)
011001	2	4.01	−0.14 (0.51)	1.26 (0.57)	−0.12 (0.51)	1.21 (0.59)
111001	1	2.57	−0.60 (0.44)	0.99 (0.69)	−0.64 (0.51)	0.86 (0.91)
000101	0	0.20	−0.40 (0.47)	0.54 (0.73)	−0.39 (0.47)	0.48 (0.76)
100101	1	0.89	−1.23 (0.62)	−0.19 (0.88)	−1.17 (0.56)	−0.17 (0.94)
010101	1	1.86	0.30 (0.58)	0.96 (0.46)	0.30 (0.60)	0.91 (0.49)
110101	0	0.67	−0.42 (0.47)	0.76 (0.70)	−0.50 (0.59)	0.50 (1.06)
001101	0	0.09	−0.39 (0.44)	0.82 (0.63)	−0.36 (0.46)	0.79 (0.67)
101101	1	0.92	−1.70 (0.43)	−0.80 (0.60)	−1.54 (0.62)	−0.62 (0.91)
011101	3	2.46	0.51 (0.56)	1.27 (0.41)	0.52 (0.58)	1.24 (0.42)
111101	2	0.59	−0.01 (0.52)	1.31 (0.54)	−0.21 (0.73)	0.90 (1.16)
000011	0	2.01	−0.90 (0.40)	1.46 (0.75)	−0.87 (0.48)	1.34 (0.82)
100011	10	9.65	−1.55 (0.77)	0.27 (0.88)	−1.47 (0.65)	0.37 (0.86)
010011	11	9.17	−0.18 (0.51)	1.37 (0.60)	−0.16 (0.51)	1.31 (0.60)
110011	3	5.07	−0.58 (0.43)	1.13 (0.67)	−0.61 (0.47)	1.02 (0.80)
001011	2	0.83	−0.68 (0.40)	1.34 (0.67)	−0.66 (0.44)	1.31 (0.70)
101011	4	8.90	−2.01 (0.51)	−0.43 (0.52)	−1.90 (0.65)	−0.27 (0.73)
011011	11	12.10	0.09 (0.48)	1.59 (0.53)	0.10 (0.49)	1.58 (0.53)
111011	5	4.33	−0.23 (0.47)	1.53 (0.60)	−0.36 (0.61)	1.29 (0.97)

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Table 6. (*Continued*)

Pattern	Observed frequency	Expected frequency	Factor scores using ML estimates		Factor scores using multiple imputation	
			$\hat{z}_1(se(\hat{z}_1))$	$\hat{z}_2(se(\hat{z}_2))$	$\hat{z}_1(se(\hat{z}_1))$	$\hat{z}_2(se(\hat{z}_2))$
000111	0	0.18	−0.40 (0.45)	0.95 (0.63)	−0.38 (0.46)	0.90 (0.65)
100111	2	1.25	−1.71 (0.45)	−0.69 (0.59)	−1.57 (0.60)	−0.53 (0.83)
010111	8	5.48	0.48 (0.55)	1.33 (0.42)	0.49 (0.57)	1.30 (0.43)
110111	2	1.11	0.01 (0.49)	1.40 (0.52)	−0.10 (0.61)	1.18 (0.86)
001111	0	0.12	−0.19 (0.48)	1.25 (0.55)	−0.17 (0.49)	1.23 (0.58)
101111	7	4.91	−2.06 (0.44)	−1.13 (0.54)	−1.94 (0.69)	−0.90 (0.89)
011111	30	29.03	0.86 (0.64)	1.62 (0.43)	0.87 (0.64)	1.63 (0.44)
111111	3	4.80	0.44 (0.48)	1.84 (0.51)	0.36 (0.69)	1.68 (0.93)

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

approximation can produce fast and accurate solutions using even only two points on each dimension. Alternatively, Laplace approximations (see Huber, Ronchetti, & Victoria-Feser, 2004) and Monte Carlo methods can be used. These approximations are expected to work better as the number of items increases, since the integrand tends to be proportional to a Gaussian. We should note that Laplace approximations succeed for continuous as well as count and binomial data. However, a small number of dichotomous items can lead to a severe integration error.

Furthermore, a method to obtain factor scores is proposed that explicitly approximates the modes of the integrated posterior distribution $p(\boldsymbol{z}|y)$ taking into account the variability of the ML estimates This method will be particularly useful for small sample sizes since the variability around the ML estimates would be greater. However, these attractive features are achieved at the expense of L extra optimizations.

Finally, some forms of the proposed GLV model can be fitted using the freet available ltm package (Rizopoulos, 2006) for R available from http://cran.r-project.org/.

References

Arminger, G., & Muthén, B. (1998). A Bayesian approach to nonlinear latent variables models using the Gibbs sampler and the Metropolis–Hastings algorithm. *Psychometrika*, *63*, 271–300.

Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology*, *34*, 93–99.

Bartholomew, D. J. (1985). Foundations of factor analysis: Some practical implications. *British Journal of Mathematical and Statistical Psychology*, *38*, 1–10.

Bartholomew, D. J. (1998). Scaling unobservable constructs in social science. *Applied Statistics*, *47*, 1–13.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London: Arnold.

Bartholomew, D. J., & McDonald, R. P. (1986). The foundations of factor analysis: A further comment. *British Journal of Mathematical and Statistical Psychology*, *39*, 228–229.

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. (2008). *Analysis of multivariate social science data*. Boca Raton, FL: Chapman & Hall/CRC.

Bartlett, M. S. (1953). Factor analysis in psychology as a statistician sees it. *Uppsala Symposium Psychological Factor Analysis* (pp. 23–34). Nordisk Psykologi's Monograph Series No.3. Copenhagen: Ejnav Mundsgaards Stockholm: Almqvist & Wiksell.

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

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

Bollen, K. (1995). Structural equation models that are nonlinear in latent variables: A least-squares estimator. *Sociological Methodology*, *25*, 223–251.

Bollen, K. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equation. *Psychometrika*, *61*, 109–121.

Bollen, K., & Paxton, P. (1998). Two-stage least squares estimation of interaction effects. In R. Schumacker & G. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation models* (pp. 125–151). Mahwah, NJ: Erlbaum.

Booth, J., & Hobert, J. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B*, *61*, 265–285.

Busemeyer, J., & Jones, L. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, *93*, 549–562.

Cox, D., & Hinkley, D. (1974). *Theoretical statistics*. London: Chapman & Hall.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Etezadi-Amoli, J., & McDonald, R. P. (1983). A second generation nonlinear factor analysis. *Psychometrika*, *48*, 315–342.

Gibson, W. (1960). Nonlinear factors in two dimensions. *Psychometrika*, *25*, 381–392.

Huber, P., Ronchetti, E., & Victoria-Feser, M. (2004). Estimation of generalized latent trait models. *Journal of the Royal Statistical Society, Series B*, *66*, 893–908.

Jöreskog, K. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, *24*, 387–404.

Jöreskog, K. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*, 381–389.

Jöreskog, K., & Yang, F. (1996). Nonlinear structural equation models: The Kenny–Judd model with interaction effects. In G. Marcoulides & R. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 57–88). Mahwah, NJ: Erlbaum.

Kenny, D., & Judd, C. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.

Lange, K. (2004). *Optimization*. New York: Springer-Verlag.

Lawley, D., & Maxwell, A. (1971). *Factor analysis as a statistical method* (2nd ed.). London: Butterworth.

Lee, S.-Y., Poon, W.-Y., & Bentler, P. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika*, *57*, 89–105.

Lee, S.-Y., & Song, X.-Y. (2004a). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, *57*, 131–150.

Lee, S.-Y., & Song, X.-Y. (2004b). Maximum likelihood analysis of a generalized latent variable model with hierarchically mixed data. *Biometrics*, *60*, 624–636.

Lee, S.-Y., Song, X.-Y., & Poon, W. -Y. (2004). Comparison of approaches in estimating interaction and quadratic effects of latent variables. *Multivariate Behavioral Research*, *39*, 37–67.

Lee, S.-Y., & Zhu, H.-T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika*, *67*, 189–210.

Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: An application. *Applied Statistics*, *50*, 325–335.

McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

McDonald, R. P. (1962). A general approach to non-linear factor analysis. *Psychometrika*, *27*, 397–415.

McDonald, R. P. (1967a). Factor interaction in nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology*, *20*, 205–215.

McDonald, R. P. (1967b). *Non-linear factor analysis*. Psychometric Monograph, 15. Grensbaro, NC: Psychonetic society.

McDonald, R. P. (1967c). Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika*, *32*(1), 77–111.

Moulder, B., & Algina, J. (2002). Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling*, *9*, 1–19.

Moustaki, I. (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology*, *56*, 337–357.

Moustaki, I., & Knott, M. (2000). Generalized latent trait models. *Psychometrika*, *65*, 391–411.

Moustaki, I., & Steele, F. (2005). Latent variable models for mixed categorical and survival responses, with an application to fertility preferences and family planning in Bangladesh. *Statistical Modelling*, *5*, 327–342.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Muthén, B. (1984). A general structural model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.

Ping, R. (2005). *Latent variable interactions and quadratics in survey data: A source book for theoretical model testing* (2nd ed.) On-line monograph, http://home.att.net/~rpingjr/intquad2/toc2.htm, last accessed March 2007.

Pinheiro, J., & Bates, D. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, *4*, 12–35.

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, *17*(5), 1–25.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Schilling, R., & Bock, S. G. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive numerical integration. *Psychometrika*, *70*, 533–555.

Schumacker, R., & Marcoulides, G. (1998). *Interaction and nonlinear effects in structural equation models*. Mahwah, NJ: Erlbaum.

Shi, J., & Lee, S.-Y. (1998). Bayesian sampling-based approach for factor analysis model with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, *51*, 233–252.

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

Song, X.-Y., & Lee, S.-Y. (2004). Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, *57*, 29–52.

Song, X.-Y., & Lee, S.-Y. (2006). Bayesian analysis of latent variable models with non-ignorable missing outcomes from exponential family. *Statistics in Medicine*, *26*, 681–693.

Wall, M., & Amemiya, Y. (2000). Estimation of polynomial structural equation analysis. *Journal of the American Statistical Association*, *95*, 929–940.

Wall, M., & Amemiya, Y. (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics*, *26*, 1–29.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrics*, *50*, 1–25.

Yang, F. (1997). *Non-linear structural equation models: Simulation studies of the Kenny-Judd model*. Unpublished doctoral dissertation, Uppsala University.

Zhu, H.-T., & Lee, S.-Y. (1999). Statistical analysis of nonlinear factor analysis models. *British Journal of Mathematical and Statistical Psychology*, *52*, 225–242.

Received 8 April 2006; revised version received 26 April 2007