Estimation of Generalized Linear Latent Variable Models

Huber, Philippe; Ronchetti, Elvezio; Victoria-Feser, Maria-Pia

How to cite

HUBER, Philippe, RONCHETTI, Elvezio, VICTORIA-FESER, Maria-Pia. Estimation of Generalized Linear Latent Variable Models. In: Journal of the Royal Statistical Society. Series B, Statistical methodology, 2004, vol. 66, n° 4, p. 893–908. doi: 10.1111/j.1467-9868.2004.05627.x

This publication URL: https://archive-ouverte.unige.ch/unige:6458

Publication DOI: 10.1111/j.1467-9868.2004.05627.x

Estimation of Generalized Linear Latent Variable Models

Philippe Huber*,§ Elvezio Ronchetti,* Maria-Pia Victoria-Feser¶,§

June 2003
Revised March 2004

Abstract

Generalized Linear Latent Variable Models (GLLVM), as defined in Bartholomew and Knott (1999) enable modelling of relationships between manifest and latent variables. They extend structural equation modelling techniques, which are powerful tools in the social sciences. However, because of the complexity of the log-likelihood function of a GLLVM, an approximation such as numerical integration must be used for inference. This can limit drastically the number of variables in the model and lead to biased estimators. In this paper, we propose a new estimator for the parameters of a GLLVM, based on a Laplace approximation to the likelihood function and which can be computed even for models with a large number of variables. The new estimator can be viewed as a M-estimator, leading to readily available asymptotic properties and correct inference. A simulation study shows its excellent finite sample properties, in particular when compared with a well established approach such as LISREL. A real data example on the measurement of wealth for the computation of multidimentional inequality is analysed to highlight the importance of the methodology.

*Department of Econometrics, University of Geneva, Blv. Pont d'Arve 40, CH-1211 Geneva, Switzerland

¶HEC, Faculty of Economic and Social Sciences and Faculty of Psychology and Education, University of Geneva, Blv. Pont d'Arve 40, CH-1211 Geneva, Switzerland

§Partially supported by Swiss National Science Foundation, grant 610-057883.99

1 Introduction

In many scientific fields researchers use models based on theoretical concepts that cannot be observed directly. This is particularly the case in social sciences. In economics, for example, there is a vast literature on welfare (see e.g. Sen, 1987) which involves measuring the standard of living of people or households in different economies. In psychology, researchers often use theoretical concepts such as intelligence, anxiety, etc, which are very important within the framework of theoretical models. However, when these models are validated by means of observed data, the problem of measurement arises. How can welfare or intelligence be measured? For welfare, income is often taken as a substitute, and in psychology, researchers have developed a battery of tests to measure intelligence indirectly.

In these situations, the researcher deals with theoretical concepts that are not observable directly (they are latent) and on the other hand, to validate the models, he (or she) uses observable quantities (manifest variables) that are proxies for the concepts of interest. This problem is not new and statistical methods have long been available; see e.g. Jöreskog (1969), Bartholomew (1984a), and Arminger and Küsters (1988). Factor analysis is one such model. A model is proposed to link manifest variables (supposed to be multivariate normal) with latent variables (or factors) and likelihood analysis can be carried out. Since the work of Jöreskog (1969), much research has been done to extend simple factor analysis to more constrained models under the heading of covariance structure or structural equations modelling. Most of these developments are readily available in software, such as LISREL (see Jöreskog, 1990; Jöreskog and Sörbom, 1993) or `gllamm` in the package `Stata` (Rabe-Hesketh, Skrondal and Pickles, 2004).

Although LISREL incorporates methods dealing with a wide variety of applied problems, it assumes that the manifest variables are multivariate normal. When this is obviously not the case (as in the case of binary variables), the manifest variables are taken as indirect observations of multivariate normal variables.

In our opinion, it is essential that the manifest variables are treated as they are, i.e. binary, ordinal or continuous, and that the model that formalizes the relationship between the manifest and the latent variables should take the type of data into account. Such models were first investigated by Bartholomew (1984a,b) who considered the case of binary data. More recently, Moustaki (1996) and Moustaki and Knott (2000) considered mixtures of manifest variables. They proposed a generalized linear latent variable model (GLLVM) (Bartholomew and Knott, 1999) that allows one to link latent variables to manifest variables of different type (see Section 2.2).

The statistical analysis of GLLVM presents a difficulty: since the latent variables are not observed, they must be integrated out from the likelihood function. One could consider several approaches to solve this problem. Moustaki (1996) and Moustaki and Knott (2000) propose using a simple Gauss - Hermite quadrature as a numerical approximation method. While this is feasible in fairly simple models, its application is often infeasible when the number of latent variables is larger than two; see Section 5. Moreover, it is known that a simple Gauss - Hermite quadrature can completely miss the maximum for certain functions and can be inefficient in other cases.

A possible improvement is provided by an adaptive Gaussian quadrature which appropriately centers and rescales the quadrature nodes and consequently is much less likely to miss the maximum; it requires a many fewer quadrature points (Rabe-Hesketh, Skrondal and Pickles, 2002). This technique is implemented in the function `gllamm` in `Stata` to fit generalized latent and mixed models and can be used to fit our models. However, due to a long computing time, the resulting estimators could not be compared in Section 5.

We propose instead using the Laplace approximation of the likelihood function. This idea has been used in other models as for example by Davison (1992). In the case of generalized linear mixed models (GLLAMM), which can be seen as a generalization of GLLVM, a simplified version of the Laplace approximation is used by Breslow and Clayton (1993) and Lin and Breslow (1996) which results in the same estimator as that proposed by McGilchrist (1994) and Lee and Nelder (1996) (see also Section 3.3). Laplace approximation of the likelihood has the important advantage with respect to quadrature that it allows one to estimate more complex models as well as models with correlated latent variables. Moreover, a direct estimation of individual scores on the latent variables space (see Section 3.3) and the statistical properties of the estimator to carry out valid inference can be easily derived. Finally, alternative estimation methods include methods based on stochastic approximations such as MCMC and MCEM; see Yau and McGilchrist (1996). While these methods have been applied successfully in many complex situations, there are potential drawbacks such as long computation times and stopping rules.

The paper is organized as follows: in Section 2, we briefly introduce the underlying variable approach implemented in e.g. LISREL used to deal with non normal manifest variables and the GLLVM. In Section 3, we propose a new estimator for the GLLVM based on the Laplace approximation of the likelihood function, investigate its statistical properties and compare it to similar estimators. The explicit formulae in the case of a GLLVM with binomial and a mixture of normal and binomial manifest variables are given in the Appendix. In Section 4, we show that the model has multiple solutions and a procedure is proposed to constrain the solution to be unique. We compare our estimator with those provided by LISREL and the GLLVM with the Gauss - Hermite quadrature in Section 5. This reveals that the new estimator has better performance in terms of bias and variance. In Section 6 we analyse a real data set from a consumption survey in Switzerland to build wealth indicators to be used for inequality measurement.

2 Two approaches for modelling latent variables

2.1 The underlying variable approach of LISREL

The underlying variable approach assumes that all the manifest variables are multivariate normal. If a variable is not normal, it is assumed to be an indirect observation of an underlying normal variable. This approach can be formulated as follows. Let X be a Bernoulli manifest variable, $\mathbf{z}$ a vector of latent variables, and $\boldsymbol{\alpha}$ a matrix of parameters. Let the conditional distribution of Y given $\mathbf{z}$ be normal with mean $\boldsymbol{\alpha}^T\mathbf{z}$ and unit variance. Given $\mathbf{z}$, a link is then established between X and Y in that it is assumed that X takes the value 1 if Y is positive and 0 otherwise. Then,

$$E(X|\mathbf{z}) = P(Y > 0 \mid \mathbf{z}) = \Phi(\boldsymbol{\alpha}^T\mathbf{z}),$$

where $\Phi(\cdot)$ is the normal cumulative distribution function. We obtain from the last equation that

$$\text{probit}\{E(X \mid \mathbf{z})\} = \Phi^{-1}\{E(X \mid \mathbf{z})\} = \boldsymbol{\alpha}^T\mathbf{z}.$$

Consequently, the assumption of an underlying normal variable in the LISREL approach can be compared to the one with the GLLVM (see below), except that the link function is a probit instead of a logit. These two link functions are very close (Lord and Novick, 1968), so that in our simulations the estimators provided by LISREL can be compared to the ones we propose in this paper (see Section 5).

In practice, the model parameters are estimated in three steps (Jöreskog, 1969, 1990). First, the thresholds of the underlying variables are estimated from the univariate means of the manifest variables. In a second step, the correlation matrix between manifest and underlying variables is estimated using polychoric, polyserial and Pearson correlations and, finally, the model parameters are obtained from a factor analysis.

2.2 Generalized Linear Latent Variable Model

In this section, we present the GLLVM starting from the framework of generalized linear models (McCullagh and Nelder, 1989). The purpose of a GLLVM is to describe the relationship between p manifest variables $x^{(j)}$, $j = 1, \ldots, p$, and $q < p$ latent variables $z^{(k)}$, $k = 1, \ldots, q$. It is assumed that the latent variables explain the observed responses in the manifest variables, so that the underlying distribution functions are the conditional distributions $g_j(x^{(j)}|\mathbf{z})$, $j = 1, \ldots, p$, which belong to the exponential family

$$g_j(x^{(j)}|\mathbf{z}) = \exp\left\{(x^{(j)}\boldsymbol{\alpha}_j^T\mathbf{z} - b_j(\boldsymbol{\alpha}_j^T\mathbf{z}))/\phi_j + c_j(x^{(j)}, \phi_j)\right\} \tag{1}$$

and $\mathbf{z} = [1, z_1, \ldots, z_q]^T = [1, \mathbf{z}_{(2)}^T]^T$. Each distribution g_j will then depend on the type of manifest variable $x^{(j)}$, as well as on a set of parameters $\boldsymbol{\alpha}_j = [\alpha_{j0}, \ldots, \alpha_{jq}]^T$ (also called loadings) and scale ϕ_j.

The essential assumption in GLLVM is that, given the latent variables, the manifest variables are conditionally independent. In other words, the latent variables explain all the dependence structure between the manifest variables. Hence, the joint conditional distribution of the manifest variable is given by $\prod_{j=1}^p g_j(x^{(j)}|\mathbf{z})$. It is also assumed that the density $h(\mathbf{z}_{(2)})$ of the latent variables is standard normal and that they are independent. This last assumption can be relaxed (Section 3). The joint distribution of the manifest and latent variables is

$$\prod_{j=1}^{p} g_j(x^{(j)}|\mathbf{z})h(\mathbf{z}_{(2)}). \tag{2}$$

Since the latent variables $\mathbf{z}_{(2)}$ are not observed, their realizations are treated as missing, and are integrated out. One actually considers the marginal density of the manifest variables

$$f_{\boldsymbol{\alpha},\boldsymbol{\phi}}(\mathbf{x}) = \int \left\{\prod_{j=1}^{p} g_j(x^{(j)}|\mathbf{z})\right\} h(\mathbf{z}_{(2)})\mathbf{dz}_{(\mathbf{2})}. \tag{3}$$

Note that $g_j(x^{(j)}|\mathbf{z})$ may be either normal or binomial according to j, or, indeed, another exponential family distribution. Our aim is to obtain estimators for the parameters $\boldsymbol{\alpha}_j$

and ϕ_j, with $j = 1 \ldots p$. Then one can use them to establish a relationship between the manifest variables $\mathbf{x}$ and the latent variables $\mathbf{z}$.

Note also that (3) formulates the general approach used with missing values (Dempster, Laird and Rubin, 1977). However, an explicit expression for (3) avoiding the integration doesn't exist, and a numerical approximation is needed. The EM algorithm can then be used to find the approximate maximum likelihood estimator of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$, as is for example pointed out in Sammel, Ryan and Legler (1997). Notice that a numerical approximation is performed within each step of the EM algorithm.

Let us now consider a sample, $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i = [x_i^{(1)}, \ldots, x_i^{(p)}]$, $i = 1, \ldots, n$. Let $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p]$ be a $(q+1) \times p$ matrix of structural parameters, and $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_p]$ the vector of scale parameters. Then the log-likelihood is

$$l(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x}) = \sum_{i=1}^{n} \log \int \left[\prod_{j=1}^{p} \exp\left\{ \frac{x_i^{(j)} \boldsymbol{\alpha}_j^T \mathbf{z} - b_j(\boldsymbol{\alpha}_j^T \mathbf{z})}{\phi_j} + c_j(x_i^{(j)}, \phi_j) \right\} \right] h(\mathbf{z}_{(2)}) \mathbf{dz}_{(\mathbf{2})} \qquad (4)$$

where b_j and c_j are known functions that depend on the chosen distribution g_j (McCullagh and Nelder, 1989).

Equation (4) contains a multidimensional integral which cannot be computed explicitly, except when all the $g_j(x^{(j)}|\mathbf{z})$ are normal. Consequently, an approximation of this integral is needed, on which the bias and variance of resulting estimators will depend.

3 Estimators based on Laplace approximation

The Gauss - Hermite quadrature (GHQ) approximation to the integral in (4) proposed by Moustaki (1996) is easy to implement but has several drawbacks. Firstly, the accuracy increases with the number of quadrature points, but decreases exponentially with the number of latent variables q. As a consequence, it is impossible in practice to handle more than two latent variables. Secondly, making correct inference on the resulting estimators seems to be very difficult. Finally, the resulting estimator appears to be biased; see Section

5.

With the Laplace approximation, inference is easier and the error rate is of order p^{-1}, where p is the number of manifest variables. This property means that the approximation improves as the number of latent variables grows (because with more latent variables one needs more manifest variables). The Laplace approximation is also well designed for functions with a single optimum, which is the case of our likelihood function. In addition, the Laplace approximation yields automatically estimates of individual scores $\hat{\mathbf{z}}_{i(2)}$ on the latent variable space; see Section 3.3. Finally, in our simulations, we found that it leads to approximately unbiased estimators; see Section 5.

3.1 Approximation of the likelihood function

By (1) and (3), the marginal distribution $f_{\boldsymbol{\alpha},\boldsymbol{\phi}}(\mathbf{x})$ can be written as

$$f_{\boldsymbol{\alpha},\boldsymbol{\phi}}(\mathbf{x}_i) = \int e^{p\cdot Q(\boldsymbol{\alpha},\boldsymbol{\phi},\mathbf{z},\mathbf{x}_i)} \mathbf{dz_{(2)}}, \tag{5}$$

where

$$Q(\boldsymbol{\alpha},\boldsymbol{\phi},\mathbf{z},\mathbf{x}_i) = \frac{1}{p}\left[\sum_{j=1}^{p}\left\{\frac{x_i^{(j)}\boldsymbol{\alpha}_j^T\mathbf{z} - b_j(\boldsymbol{\alpha}_j^T\mathbf{z})}{\phi_j} + c_j(x_i^{(j)},\phi_j)\right\} - \frac{\mathbf{z}_{(2)}^T\mathbf{z}_{(2)}}{2} - \frac{q}{2}\log(2\pi)\right]. \tag{6}$$

Applying the q-dimensional Laplace approximation to the density (5) (De Bruijn, 1981; Tierney and Kadane, 1986), we obtain

$$f_{\boldsymbol{\alpha},\boldsymbol{\phi}}(\mathbf{x}_i) = \left(\frac{2\pi}{p}\right)^{q/2}\left[\det\{-\mathbf{U}(\hat{\mathbf{z}}_i)\}\right]^{-1/2} e^{pQ(\boldsymbol{\alpha},\boldsymbol{\phi},\hat{\mathbf{z}}_i,\mathbf{x}_i)}\{1+O(p^{-1})\}, \tag{7}$$

where

$$\mathbf{U}(\hat{\mathbf{z}}_i) = \left.\frac{\partial^2 Q(\boldsymbol{\alpha},\boldsymbol{\phi},\mathbf{z},\mathbf{x}_i)}{\partial\mathbf{z}^T\partial\mathbf{z}}\right|_{\mathbf{z}=\hat{\mathbf{z}}_i} = -\frac{1}{p}\boldsymbol{\Gamma}(\boldsymbol{\alpha},\boldsymbol{\phi},\hat{\mathbf{z}}_i) \tag{8}$$

$$\boldsymbol{\Gamma}(\boldsymbol{\alpha},\boldsymbol{\phi},\hat{\mathbf{z}}_i) = \sum_{j=1}^{p}\frac{1}{\phi_j}\left.\frac{\partial^2 b_j(\boldsymbol{\alpha}_j^T\mathbf{z})}{\partial\mathbf{z}^T\partial\mathbf{z}}\right|_{\mathbf{z}=\hat{\mathbf{z}}_i} + \mathbf{I}_q, \tag{9}$$

and $\hat{\mathbf{z}}_i = [1\ \hat{\mathbf{z}}_{i(2)}]$ is the maximum of $Q(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{z}, \mathbf{x}_i)$, i.e. the root of $\partial Q(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{z}, \mathbf{x}_i)/\partial \mathbf{z} = 0$ defined through the iterative equation

$$\hat{\mathbf{z}}_{i(2)} = \hat{\mathbf{z}}_{i(2)}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{x}_i) = \sum_{j=1}^{p} \frac{1}{\phi_j} \left\{ x_i^{(j)} - \frac{\partial b_j(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)}{\partial \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i} \right\} \boldsymbol{\alpha}_{j(2)}, \quad i = 1, \ldots, n, \tag{10}$$

with $\boldsymbol{\alpha}_j = [\alpha_{j0}, \boldsymbol{\alpha}_{j(2)}^T]^T$.

Notice that there are n vectors $\hat{\mathbf{z}}_{i(2)}$ to be determined by the implicit equations (10) and each $\hat{\mathbf{z}}_{i(2)}$ depends on all the parameters of the model and the observation $\mathbf{x}_i$.

3.2 Laplace approximated maximum likelihood estimators

The Laplace approximation eliminates the integral from the marginal distribution of $\mathbf{x}_i$. From (6), (7), (9), and (9), we obtain the approximate log-likelihood function

$$\begin{aligned}\tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x}) = \sum_{i=1}^{n} \Bigg(& -\frac{1}{2} \log\ \det\{\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)\} \\ & + \sum_{j=1}^{p} \left\{ \frac{x_i^{(j)} \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i - b_j(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)}{\phi_j} + c_j(x_i^{(j)}, \phi_j) \right\} - \frac{\hat{\mathbf{z}}_{i(2)}^T \hat{\mathbf{z}}_{i(2)}}{2} \Bigg).\end{aligned} \tag{11}$$

The new estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ based on the Laplace approximation are found by equating the derivative of (11) to zero and inserting (10) into (11). For the structural parameters $\boldsymbol{\alpha}$, we have

$$\frac{\partial \tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x})}{\partial \alpha_{kl}} = \sum_{i=1}^{n} \left[-\frac{1}{2} \mathrm{tr} \left\{ \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)^{-1} \frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)}{\partial \alpha_{kl}} \right\} + \frac{1}{\phi_k} \left\{ x_i^{(k)} - \left. \frac{\partial b_k(\boldsymbol{\alpha}_k^T \mathbf{z})}{\partial \boldsymbol{\alpha}_k^T \mathbf{z}} \right|_{\mathbf{z}=\hat{\mathbf{z}}_i} \right\} \hat{z}_{il} \right] = 0, \tag{12}$$

where $\hat{z}_{il}$ is the l^{th} element of the vector $\hat{\mathbf{z}}_i$ and $\frac{\partial \boldsymbol{\Gamma}}{\partial \boldsymbol{\alpha}_{kl}}$ is the $(q \times q)$ matrix obtained from $\boldsymbol{\Gamma}$ by differentiating all its elements with respect to α_{kl}, $k = 1, \ldots, p$, $l = 0, \ldots, q$.

Similarly, for $\boldsymbol{\phi}$, we obtain

$$\begin{aligned}\frac{\partial \tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x})}{\partial \phi_k} = \sum_{i=1}^{n} \Bigg[& -\frac{1}{2} \mathrm{tr} \left\{ \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)^{-1} \frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)}{\partial \phi_k} \right\} \\ & - \frac{1}{\phi_k^2} \{ x_i^{(k)} \boldsymbol{\alpha}_k^T \hat{\mathbf{z}}_i + b_i(\boldsymbol{\alpha}_k^T \hat{\mathbf{z}}_i) \} + \frac{\partial c_k(x_i^{(j)}, \phi_k)}{\partial \phi_k} \Bigg] = 0, \quad k = 1, \ldots, p.\end{aligned} \tag{13}$$

Hence (12) and (13) provide a set of estimating equations defining the estimators for the model parameters. In addition, (10) is required for the computation of all nq terms $\mathbf{z}_{i(2)}$.

In the derivation of the estimating equations, the model has been kept as general as possible without specifying the conditional distributions $g_j(x^{(j)}|\mathbf{z})$. In the Appendix, we give specific expressions for the quantities used in the log-likelihood (11), the score functions (12) and (13), and $\hat{\mathbf{z}}_{i(2)}$ in (10) for binomial and a mixture of binomial and normal manifest variables. The analytic computations are tedious but straightforward. An alternative approach would be to use numerical derivatives in (12) and (13). In this paper, we focus our examples on binomial distributions and a mixture of normal and binomial distributions.

3.3 Interpretation of the new estimator

A way to interpret the estimators derived in Section 3.2 is to consider the $\mathbf{z}_i$ as "parameter" in (2). Then the "likelihood" would be

$$
\begin{aligned}
l^*(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{z}|\mathbf{x}) &= \sum_{i=1}^{n} \log \left\{ \prod_{j=1}^{p} g_j(x_i^{(j)}|\mathbf{z}_i) h(\mathbf{z}_{i(2)}) \right\} \\
&= \sum_{i=1}^{n} \Bigg[\sum_{j=1}^{p} \left\{ \frac{x_i^{(j)} \boldsymbol{\alpha}_j^T \mathbf{z}_i - b_j(\boldsymbol{\alpha}_j^T \mathbf{z}_i)}{\phi_j} + c_j(x_i^{(j)}, \phi_j) \right\} \\
&\qquad - \frac{\mathbf{z}_{i(2)}^T \mathbf{z}_{i(2)}}{2} - \frac{q}{2} \log(2\pi) \Bigg] = p \sum_{i=1}^{n} Q(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{z}_i, \mathbf{x}_i)
\end{aligned}
\tag{14}
$$

which differs from (11) by the additive term

$$
-\frac{1}{2} \sum_{i=1}^{n} \log\det \{ \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{z}_i) \} - n\frac{q}{2} \log(2\pi). \tag{15}
$$

Taking the derivative of l^* with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ doesn't lead to the same expressions for the score function as (12) and (13) and hence, the corresponding estimators are different.

However, taking the derivative of l^* with respect to $\mathbf{z}_{i(2)}$ leads to the same implicit equation (10) defining the $\hat{\mathbf{z}}_{i(2)}$ needed by the Laplace approximation. Hence, the $\hat{\mathbf{z}}_{i(2)}$ are directly interpretable as the "maximum likelihood estimators" of the individual latent scores. They can then be used for example to represent graphically the subject in the latent variable space.

It should be stressed that (14) defines the Lee and Nelder (1996) h-likelihood. In the context of generalized linear mixed models, (14) is recognized by Breslow and Clayton (1993, equation (6)) as Green (1987)'s penalized quasi likelihood. The maximization of the h-likelihood is also called the BLUP (best linear unbiased prediction) by McGilchrist (1994). It is then clear that because of the inclusion of the additive term (15) in the log-likelihood, the Laplace Approximated Maximum Likelihood Estimators (LAMLE) for $\boldsymbol{\alpha}, \boldsymbol{\phi}$ we propose is different from the penalized quasi likelihood estimator or indeed the maximum h-likelihood estimator. On the other hand, the estimated latent scores $\hat{\mathbf{z}}_{i(2)}$ are penalized quasi likelihood or maximum h-likelihood estimators.

3.4 Statistical properties of the Laplace approximated likelihood estimator

Let $\hat{\boldsymbol{\theta}}_L$ be the vector containing all the LAMLE of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ for a GLLVM. It is defined by the estimating equations (12) and (13), where the $\hat{\mathbf{z}}_{i(2)}$ are defined by (10).

The LAMLE $\hat{\boldsymbol{\theta}}_L$ belongs to the class of M-estimators (Huber, 1964, 1981) which are implicitly defined through a general Ψ-function as the solution in $\boldsymbol{\theta}$ of

$$\sum_{i=1}^{n} \Psi(x_i; \boldsymbol{\theta}) = 0.$$

The Ψ-function for the LAMLE is given by (12) and (13). Then, under the conditions given in Huber (1981, pp. 131 - 133) or Welsh (1996, p. 194), the LAMLE is consistent and asymptotically normal, i.e.

$$n^{1/2}(\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N(0, B(\boldsymbol{\theta}_0)^{-1} A(\boldsymbol{\theta}_0) B(\boldsymbol{\theta}_0)^{-T})$$

as $n \to \infty$, where

$$A(\boldsymbol{\theta}_0) = E\left[\Psi(x;\boldsymbol{\theta}_0)\Psi^T(x;\boldsymbol{\theta}_0)\right], \quad B(\boldsymbol{\theta}_0) = -E\left[\frac{\partial\Psi(x;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right].$$

These conditions must be checked for each particular conditional distribution g_j.

Moreover, the function $\tilde{l}(\boldsymbol{\alpha},\boldsymbol{\phi}|\mathbf{x})$ in (11) plays the role of a pseudo-likelihood function and can be used to construct likelihood-ratio type tests as in Heritier and Ronchetti (1994, p. 898), by defining $\rho(x;\boldsymbol{\theta}) = -\tilde{l}(\boldsymbol{\alpha},\boldsymbol{\phi}|\mathbf{x})$. This allows one to carry out inference and variable selection in GLLVM.

4 Constraints and correlated latent variables

The estimating equations which define the LAMLE, or the maximum likelihood estimator, may have multiple solutions. In this section, we first investigate the number of constraints which are required to make the solution unique and we propose a procedure to select those constraints. Then, we extend the LAMLE to the case of correlated latent variables.

4.1 Constraining the Laplace approximated likelihood estimators

Let us recall that the GLLVM model is based upon a generalized linear model. Therefore,

$$\nu(E(\mathbf{x}|\mathbf{z})) = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}^T\mathbf{z}_{(2)},$$

where $\nu(\cdot)$ is a link function and we define $\mathbf{z}_{(2)}$ to be centered and standardized. Let $\mathbf{H}$ be an orthogonal matrix of dimension $q \times q$. It is possible to rotate the matrix $\boldsymbol{\alpha}$ by pre-multiplying it by $\mathbf{H}$ and to obtain a new matrix of parameters $\boldsymbol{\alpha}^* = \mathbf{H}\boldsymbol{\alpha}$. Since $\mathbf{z}_{(2)}$ is centered and standardized and $\mathbf{H}$ is orthogonal, $\mathbf{z}^*_{(2)} = \mathbf{H}\mathbf{z}_{(2)}$ is standard normal. Moreover, the rotation $\mathbf{H}$ does not change the following product:

$$\boldsymbol{\alpha}^{*T}\mathbf{z}^*_{(2)} = \boldsymbol{\alpha}^T\mathbf{H}^T\mathbf{H}\mathbf{z}_{(2)} = \boldsymbol{\alpha}^T\mathbf{z}_{(2)}.$$

Therefore, a rotation of $\boldsymbol{\alpha}$ gives a new matrix of parameters which is also a solution for the same model. This is the problem encountered in factor analysis. If a unique solution is required, it is necessary to impose constraints on the parameters $\boldsymbol{\alpha}$. Notice that this is typically done in exploratory model setting, where instead of constraining the matrix $\boldsymbol{\alpha}$, one uses, for instance, a *varimax* rotation.

An orthogonal matrix of size $q \times q$ possesses $q(q-1)/2$ degrees of freedom. In other words, such a matrix needs at least $q(q-1)/2$ constraints on its elements to be unique and this represents the number of constraints we have to impose to obtain a unique solution for the model.

Proposition *Let $\hat{\boldsymbol{\alpha}}$ be a matrix of dimension $q \times p$ containing the LAMLE of $\boldsymbol{\alpha}$. If all the elements of the upper triangle of $\hat{\boldsymbol{\alpha}}^T$ are constrained, then $\hat{\boldsymbol{\alpha}}^T$ is completely determined, except for the sign of each column. If at least one constraint of the j^{th} column, with $j = 2, \ldots, q$, is different from zero, then the sign of the corresponding column is determined.*

The proof is given in Appendix B.

4.2 Laplace approximated likelihood estimators of a generalized linear latent variable model with correlated latent variables

The flexible form of the Laplace approximation allows us to handle correlated latent variables. Let $\boldsymbol{\Sigma}$ be the correlation matrix of the latent variables and consider latent variables with unit variance. Then, the density of $\mathbf{z}_{(2)}$ becomes

$$h(\mathbf{z}_{(2)}) = (2\pi)^{-q/2} \left|\det \Sigma\right|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{z}_{(2)}^T \boldsymbol{\Sigma}^{-1} \mathbf{z}_{(2)}\right),$$

which implies that the function Q, defined by (6), is modified as follows:

$$Q(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{z}, \mathbf{x}_i, \boldsymbol{\Sigma}) = \frac{1}{p}\left[\sum_{j=1}^{p}\left\{\frac{x_i^{(j)}\boldsymbol{\alpha}_j^T\mathbf{z} - b_j(\boldsymbol{\alpha}_j^T\mathbf{z})}{\phi_j} + c_j(x_i^{(j)}, \phi_j)\right\} - \frac{\mathbf{z}_{(2)}^T\boldsymbol{\Sigma}^{-1}\mathbf{z}_{(2)}}{2} - \frac{q}{2}\log(2\pi)\right]. \tag{16}$$

Therefore, the implicit equation (10) defining $\mathbf{z}_{(2)}$ becomes

$$\hat{\mathbf{z}}_{i(2)} := \hat{\mathbf{z}}_{i(2)}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{x}_i, \boldsymbol{\Sigma}) = \sum_{j=1}^{p} \frac{1}{\phi_j} \left\{ x_i^{(j)} - \frac{\partial b_j(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)}{\partial \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i} \right\} \boldsymbol{\Sigma} \boldsymbol{\alpha}_{j(2)}. \tag{17}$$

Let

$$\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i, \boldsymbol{\Sigma}) = \sum_{j=1}^{p} \frac{1}{\phi_j} \frac{\partial^2 b_j(\boldsymbol{\alpha}^T \mathbf{z})}{\partial \mathbf{z} \partial \mathbf{z}^T} \bigg|_{\mathbf{z}=\hat{\mathbf{z}}_i} + \boldsymbol{\Sigma}^{-1}. \tag{18}$$

The estimating equations defining the LAMLE with correlated latent variables are the modified (12) and (13) using (18) and, in addition, the $q(q-1)/2$ equations for the elements σ_{kl} of $\boldsymbol{\Sigma}$:

$$\begin{aligned} \frac{\partial \tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi} | \mathbf{x})}{\partial \sigma_{kl}} = \sum_{i=1}^{n} \Bigg[& -\frac{1}{2} \mathrm{tr} \left\{ \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i, \boldsymbol{\Sigma})^{-1} \frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i, \boldsymbol{\Sigma})}{\partial \sigma_{kl}} \right\} \\ & -\frac{1}{2} \mathrm{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{kl}} \right) + \frac{1}{2} \hat{\mathbf{z}}_{i(2)}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{kl}} \boldsymbol{\Sigma}^{-1} \hat{\mathbf{z}}_{i(2)} \Bigg] = 0. \end{aligned} \tag{19}$$

5 Simulation study

In this section, we compare the LAMLE with uncorrelated latent variables with the maximum likelihood estimator using the GHQ approximation and the LISREL estimators that we take as the benchmarks. We have considered models containing one, two, and four latent variables but we present here the results only for four latent variables. The former can be found in Huber, Ronchetti and Victoria-Feser (2003). Since the GHQ approximation for more than two latent variables is not available, we perform the simulations with four latent variables only for the LAMLE and LISREL estimators.

We also tried to compare the performance of the LAMLE with the maximum likelihood estimator with adaptative GHQ approximation as implemented in the package `gllamm` in `Stata`, but as we mentioned above, it took about 20 hours to compute the approximate maximum likelihood estimator of the parameters of the simple model with 10 manifest variables, 2 latent variables and 60 observations. Therefore, we didn't pursue this comparison further.

5.1 Design

To study the behavior of the LAMLE and to compare it with the benchmarks, we generate samples from GLLVM with known parameters. As we showed in Section 2, this design can be used to compare the LAMLE with the estimates provided by LISREL because they can be interpreted as generalized linear models with a probit link function.

Random samples of size n are generated in S-Plus. The procedure is as follows:

1. Initialize all the parameters:
   - $p(q+1)$ elements in the matrix $\boldsymbol{\alpha}$,
   - p_1 variances defining the vector $\boldsymbol{\phi}$ for the normal variables.
2. Generate q independent standard normal vectors $\mathbf{z}$ of size n.
3. Generate a vector $\boldsymbol{\mu} = E[X|\mathbf{z}]$ of conditional means of all responses defined by

$$\boldsymbol{\nu}(\boldsymbol{\mu}) = \boldsymbol{\alpha}^T \mathbf{z},$$

   $\boldsymbol{\nu}$ being the link functions corresponding to the distributions of each manifest variable.
4. Generate all responses $\mathbf{x}$ based upon the means $\boldsymbol{\mu}$ that were calculated in step 3 as well as the scale parameters $\boldsymbol{\phi}$ for the normal responses.

A quasi-Newton procedure (Dennis and Schnabel, 1983) is used to solve the implicit equations (10), (12) and (13). The algorithm is written in `C` and the program is available from the authors upon request. For the LISREL estimators, the covariance matrix is computed using LISREL 8.51 and a factor analysis is then performed with S-Plus. Then, the estimators for the binomial loadings are multiplied by 1.7 to make them comparable with the LAMLE; see Section 2.1.

500 samples of size 400 were simulated. They contain 8 normal and 8 binomial responses with 4 latent variables. The parameters are given in Table 1.

	Normal items							
$\boldsymbol{\alpha}_0$	3.2	3.3	3.1	3.5	3.2	3.4	3.3	3.6
$\boldsymbol{\alpha}_1$	-2.0	-4.0	7.0	0.0	5.0	-8.0	-8.0	-3.0
$\boldsymbol{\alpha}_2$	0.0	1.0	-3.0	-2.0	0.0	3.0	4.0	5.0
$\boldsymbol{\alpha}_3$	0.0	0.0	3.0	0.0	-1.0	2.0	4.0	-9.0
$\boldsymbol{\alpha}_4$	0.0	0.0	0.0	2.0	-4.0	2.0	6.0	-4.0
$\boldsymbol{\phi}$	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

	Binomial items							
$\boldsymbol{\alpha}_0$	-0.7	0.9	0.8	0.8	0.1	0.3	0.4	-0.8
$\boldsymbol{\alpha}_1$	0.6	0.6	-0.3	-0.6	0.0	0.5	0.3	-0.1
$\boldsymbol{\alpha}_2$	0.1	0.2	-0.3	-0.2	-0.4	0.2	0.5	-0.3
$\boldsymbol{\alpha}_3$	0.0	0.8	-0.3	-0.5	0.2	0.6	0.4	-0.5
$\boldsymbol{\alpha}_4$	0.4	0.6	-0.3	-0.2	-0.7	-0.5	-0.2	-0.3

Table 1: Parameters for a model with four latent variables

5.2 Discussion of the results

Models with a single latent variable are rather simple and all methods (including the GHQ approximation) give good results: all estimators are unbiased. The results are not shown here but can be found in Huber, Ronchetti and Victoria-Feser (2003).

In models containing two latent variables, large biases appear with the GHQ approximations and the LISREL approach; again see Huber, Ronchetti and Victoria-Feser (2003). The bias with GHQ approximation is explained by the fact that it is based upon the integration on pre-specified and fixed quadrature points. With 16 and 8 quadrature points, this grid becomes coarser and it can happen that the peak of the log-likelihood is missed. On the other hand, the Laplace approximation searches for the point that is the maximum of the likelihood and approximates adaptively (i.e. for each $\mathbf{x}_i$) the function in its neighborhood. The LISREL estimators are unbiased for normal manifest variables but show large biases for some of the binomial manifest variables.

Here, we consider the results of the model with 8 normal and 8 binomial responses and

4 latent variables. Boxplots of the values of the estimators of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_3$ are presented in Figures 1 and 2. As was the case for one and two latent variables, the LAMLE are almost unbiased. On the other hand, the LISREL estimators for the loadings of binomial manifest variables are significantly biased. Similar plots were obtained for other parameter values.

Insert Figures 1 and 2 here

---

6 Consumption data analysis

In this section, we analyze a real data set and compare the results provided by the LAMLE and LISREL. The data are from the consumption survey in 1990 in Switzerland, provided by the Swiss Federal Statistical Office. This database consists of a series of wealth indicators measured on households. Some of these variables are expenditures for general categories such as food, housing, leisure, others are the ownership by the household of items such as TV, washing machine, freezer, etc. The first type of variables are actually continuous, whereas the second type are binary (with the value 1 indicating the presence of the item). This type of data are typically used for the measurement of multidimentional inequality (e.g. Maasoumi, 1986): the wealth indicators are first aggregated to independent wealth distributions on which inequality measures are then computed. We propose here to use a GLLVM to construct the aggregated measures of wealth. We have selected the following variables (the currency is the Swiss franc) from the survey: logarithm of total income (income), logarithm of expenditure for housing and energy (housing), logarithm of expenditure for clothing (clothing), logarithm of expenditure for leisure activities (leisure), presence of a freezer (freezer), presence of a dishwasher (dishwasher), presence of a washing machine (washing), presence of a (color) television (TV), presence of a video recorder (video) and presence of a car (car).

The sample size is $n = 9907$ (after removal of missing data). We fit these data using the GLLVM with six binary and four normal manifest variables and two latent variables using the LAMLE. We also estimate the same parameters using LISREL as a comparison. The parameter estimates and the standard errors for the LAMLE are presented in Table 2. Unfortunately, no standard errors for the estimates using LISREL are available.

We can interpret these results at two levels. First, when one compares the estimates provided by the LAMLE and LISREL, we find that they are very similar for the parameters of the normal items (constants, loadings and standard deviations), but they differ quite substantially for the binary items. The loadings provided by LISREL are systematically smaller which indicates a probable bias since this feature was found in the simulation study. Second, by looking at the estimates provided by the LAMLE, one can see that the first latent variable is essentially determined by the income and expenditure variables (large loadings) whereas the second latent variable is essentially determined by the ownership variables. This suggests that concerning wealth distributions, a distinction should be made between income-expenditure and capital. However, a more extensive study is needed (that includes for example more items) before final conclusions can be drawn.

7 Acknowledgements

The authors would like to thank the Joint Editor, the Associate Editor, and two referees for very useful comments which lead to a clear improvement of the manuscript.

Appendix A: LAMLE for GLLVM with binomial and a mixture of binomial and normal manifest variables

A.1 Binomial manifest variables

Let X, with possible values $0, 1, \ldots, m$, have a binomial distribution with expectation $m \cdot \pi(\mathbf{z})$. Using the canonical link function for binomial distributions, we have

$$\pi(\mathbf{z}) = \frac{\exp(\boldsymbol{\alpha}^T \mathbf{z})}{1 + \exp(\boldsymbol{\alpha}^T \mathbf{z})}.$$

The scale parameter $\phi = 1$ and the functions b and c in (1) are

$$b(\boldsymbol{\alpha}^T \mathbf{z}) = m\log\{1 + \exp(\boldsymbol{\alpha}^T \mathbf{z})\}, \tag{20a}$$

$$c(x, \phi) = c(x) = \log\binom{m}{x}, \tag{20b}$$

and

$$g(x|\mathbf{z}) = \binom{m}{x} \pi(\mathbf{z})^x \{1 - \pi(\mathbf{z})\}^{(m-x)}. \tag{21}$$

The log-likelihood for binomial responses, using the expressions in (11) is

$$\tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x}) = \sum_{i=1}^{n} \Bigg[-\frac{1}{2}\log\, \det\big\{\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i)\big\} + \sum_{j=1}^{p} \left[x_i^{(j)} \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i - m\log\{1 + \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)\} + \log\binom{m}{x_i^{(j)}} \right] - \frac{\hat{\mathbf{z}}_{i(2)}^T \hat{\mathbf{z}}_{i(2)}}{2} \Bigg], \tag{22}$$

with

$$\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i) = \sum_{j=1}^{p} m \frac{\exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)}{\{1 + \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)\}^2} \boldsymbol{\alpha}_{j(2)} \boldsymbol{\alpha}_{j(2)}^T + \mathbf{I}_q = \sum_{j=1}^{p} m\beta_{ji} \boldsymbol{\alpha}_{j(2)} \boldsymbol{\alpha}_{j(2)}^T + \mathbf{I}_q,$$

and $\beta_{ji} = \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)(1 + \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i))^{-2}$. There, $\hat{\mathbf{z}}_{i(2)}$ is the solution of the implicit equation (see (10)):

$$\hat{\mathbf{z}}_{i(2)} = \sum_{j=1}^{p} \left\{ x_i^{(j)} - m \frac{\exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)}{1 + \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)} \right\} \boldsymbol{\alpha}_{j(2)}. \tag{23}$$

To compute the score functions, we first need

$$\mathrm{tr}\left\{\boldsymbol{\Gamma}(\boldsymbol{\alpha},\hat{\mathbf{z}}_i)^{-1}\frac{\partial\boldsymbol{\Gamma}(\boldsymbol{\alpha},\hat{\mathbf{z}}_i)}{\partial\alpha_{kl}}\right\} = \mathrm{tr}\left[\left\{\sum_{j=1}^{p} m\beta_{ji}\boldsymbol{\alpha}_{j(2)}\boldsymbol{\alpha}_{j(2)}^T + \mathbf{I}_q\right\}^{-1}\right.$$
$$\left.\left\{\sum_{j=1}^{p} m\beta_{ji}\left(\frac{1-\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}\frac{\partial\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i}{\partial\alpha_{kl}}\boldsymbol{\alpha}_{j(2)}\boldsymbol{\alpha}_{j(2)}^T\right) + (1-\delta_{l0})(\mathbf{e}_l\otimes\boldsymbol{\alpha}_k^T + \mathbf{e}_l^T\otimes\boldsymbol{\alpha}_k)\right\}\right], \tag{24}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{e}_l$ is the vector of length q whose elements are zeros except the l^{th} one which is 1. Moreover,

$$\frac{\partial b_j(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{\partial\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i} = m\frac{\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}. \tag{25}$$

Finally, by means of the generalized implicit functions theorem, we differentiate $\hat{\mathbf{z}}_{i(2)}$ and obtain

$$\frac{\partial\hat{\mathbf{z}}_{i(2)}}{\partial\alpha_{k0}} = -m\beta_{ki}\boldsymbol{\Gamma}(\boldsymbol{\alpha},\hat{\mathbf{z}}_i)^{-1}\boldsymbol{\alpha}_{k(2)} \tag{26a}$$
$$\frac{\partial\hat{\mathbf{z}}_{i(2)}}{\partial\boldsymbol{\alpha}_{k(2)}} = \boldsymbol{\Gamma}(\boldsymbol{\alpha},\hat{\mathbf{z}}_i)^{-1}\left[-m\beta_{ki}\boldsymbol{\alpha}_{k(2)}\hat{\mathbf{z}}_{i(2)}^T + \left\{x_i^{(k)} - m\frac{\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}\right\}\mathbf{I}_q\right]. \tag{26b}$$

The LAMLE of a model with binomial manifest variables is completely defined by the pseudo log-likelihood (22) and its score functions (12) whose components are given by (23), (24), (25), and (26).

A.2 Mixture of binomial and normal manifest variables

In practice, a mixture model with both normal and binomial responses is more realistic than the models we presented in A.1. Let us suppose that among the p manifest variables, the first p_1 are normal and the last $p - p_1$ follow a binomial distribution. To create the approximate model, the procedure is the same as before except that all sums over j are separated into two parts, depending on whether j is related to a normal or a binomial variable. Consequently, the pseudo log-likelihood takes the following form:

$$
\begin{aligned}
\tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x}) = \sum_{i=1}^{n} \Bigg[& -\frac{1}{2}\log\, \det\big\{\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)\big\} \\
& + \sum_{j=1}^{p_1} \Bigg[\frac{\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i}{\phi_j} \Bigg(x_i^{(j)} - \frac{\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i}{2} \Bigg) - \frac{1}{2} \Bigg(\frac{\left(x_i^{(j)}\right)^2}{\phi_j} + \log(2\pi\phi_j) \Bigg) \Bigg] \\
& + \sum_{j=p_1+1}^{p} \Bigg[x_i^{(j)} \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i - m\log\{1 + \exp(\boldsymbol{\alpha}_j^T \cdot \hat{\mathbf{z}}_i)\} + \log \binom{m}{x_i^{(j)}} \Bigg] - \frac{\hat{\mathbf{z}}_{i(2)}^T \hat{\mathbf{z}}_{i(2)}}{2} \Bigg],
\end{aligned} \tag{27}
$$

where

$$
\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i) = \sum_{j=1}^{p_1} \frac{\boldsymbol{\alpha}_{j(2)} \boldsymbol{\alpha}_{j(2)}^T}{\phi_j} + \sum_{j=p_1+1}^{p} m\beta_{jk} \boldsymbol{\alpha}_{j(2)} \boldsymbol{\alpha}_{j(2)}^T + \mathbf{I}_q = \boldsymbol{\Gamma}_{\mathbf{1}}(\boldsymbol{\alpha}, \boldsymbol{\phi}) + \boldsymbol{\Gamma}_{\mathbf{2}}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i) + \mathbf{I}_q.
$$

$\hat{\mathbf{z}}_{i(2)}$ is obtained through the implicit equation:

$$
\hat{\mathbf{z}}_{i(2)} = \sum_{j=1}^{p_1} \frac{1}{\phi_j} (x_i^{(j)} - \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i) \boldsymbol{\alpha}_{j(2)} + \sum_{j=p_1+1}^{p} \Bigg\{ x_i^{(j)} - m \frac{\exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)}{1 + \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)} \Bigg\} \boldsymbol{\alpha}_{j(2)}. \tag{28}
$$

We differentiate (27) to obtain the score functions. As normal responses are present in the model, score functions for $\boldsymbol{\phi}$ are also required. The different components of equations (12) and (13) are

$$
\begin{aligned}
\frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)}{\partial \alpha_{kl}} = & (1 - \delta_{l0}) \big(\mathbf{e}_l \otimes \boldsymbol{\alpha}_i^T + \mathbf{e}_l^T \otimes \boldsymbol{\alpha}_i \big) \big(\frac{1}{\phi_k} D_1 + m\beta_{ki} D_2 \big) \\
& + \sum_{j=p_1+1}^{p} m\beta_{ji} \Bigg\{ \frac{1 - \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)}{1 + \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)} \frac{\partial \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i}{\partial \alpha_{kl}} \boldsymbol{\alpha}_{j(2)} \boldsymbol{\alpha}_{j(2)}^T \Bigg\},
\end{aligned} \tag{29a}
$$

where

$$
D_1 = \left\{ \begin{array}{lll} 1 & : & 1 \leq i \leq p_1 \\ 0 & : & p_1 < i \leq p \end{array} \right. \qquad \text{and} \qquad D_2 = \left\{ \begin{array}{lll} 0 & : & 1 \leq i \leq p_1 \\ 1 & : & p_1 < i \leq p \end{array} \right. ,
$$

and

$$
\frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)}{\partial \phi_k} = -\frac{1}{\phi_k^2} \boldsymbol{\alpha}_{k(2)} \boldsymbol{\alpha}_{k(2)}^T + \sum_{j=p_1+1}^{p} m\beta_{ji} \frac{1 - \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)}{1 + \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)} \frac{\partial \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i}{\partial \phi_k} \boldsymbol{\alpha}_{j(2)} \boldsymbol{\alpha}_{j(2)}^T. \tag{29b}
$$

Moreover,

$$
\frac{\partial \hat{\mathbf{z}}_{i(2)}}{\partial \alpha_{k0}} = \begin{cases} -\frac{1}{\phi_k} \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi})^{-1} \boldsymbol{\alpha}_{k(2)}, & \text{if } 1 \leq i \leq p_1 \\ -m\beta_{ki} \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i)^{-1} \boldsymbol{\alpha}_{k(2)}, & \text{otherwise} \end{cases} \tag{30a}
$$

$$\frac{\partial \hat{\mathbf{z}}_{i(2)}}{\partial \alpha_{kl}} = \begin{cases} \frac{1}{\phi_k}\boldsymbol{\Gamma}(\boldsymbol{\alpha},\boldsymbol{\phi})^{-1}\Big\{-\boldsymbol{\alpha}_{k(2)}\hat{\mathbf{z}}_{i(2)}^T + (x_i^{(k)} - \boldsymbol{\alpha}_k^T\hat{\mathbf{z}}_i)\mathbf{I}_q\Big\}, & \text{if } 1 \leq i \leq p_1 \\ \boldsymbol{\Gamma}(\boldsymbol{\alpha},\hat{\mathbf{z}}_i)^{-1}\Big[-m\beta_{ki}\boldsymbol{\alpha}_{k(2)}\hat{\mathbf{z}}_{i(2)}^T + \big\{x_i^{(k)} - m\frac{\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}\big\}\mathbf{I}_q\Big], & \text{otherwise} \end{cases} \tag{30b}$$

$$\frac{\partial \hat{\mathbf{z}}_{i(2)}}{\partial \phi_k} = -\boldsymbol{\Gamma}(\boldsymbol{\alpha},\boldsymbol{\phi})^{-1}\left\{\frac{1}{\phi_k^2}(x_i^{(k)} - \boldsymbol{\alpha}_i^T\hat{\mathbf{z}}_i)\boldsymbol{\alpha}_{k(2)}\right\}, \quad \text{if } 1 \leq i \leq p_1. \tag{30c}$$

Thus, the pseudo log-likelihood (27) is maximized when the score functions given by (12) and (13) are set to zero, where expressions (23), (29) and (30) are used.

Appendix B: Proof of Proposition 1

First, we establish the proposition for a square matrix $\hat{\boldsymbol{\alpha}}$.

Let $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_{ij})_{1\leq i,j\leq q}$ and $\hat{\boldsymbol{\alpha}}^* = (\hat{\alpha}^*_{ij})_{1\leq i,j\leq q}$ be two square matrices of dimension $q \times q$ and $\mathbf{H} = (h_{ij})_{1\leq i,j\leq q}$ an orthogonal matrix of dimension $q \times q$. If $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\alpha}}^*$ have the same upper triangle, and if $\hat{\boldsymbol{\alpha}} = \mathbf{H}\hat{\boldsymbol{\alpha}}^*$, then it is straightforward to show that $\mathbf{H}$ is diagonal, i.e. $h_{ij} = \pm\delta_{ij}$, with $1 \leq i, j \leq q$ and δ_{ij} the Kronecker symbol.

The extension to matrices of dimension $p\times q$ is trivial as $\hat{\boldsymbol{\alpha}}$ (resp. $\hat{\boldsymbol{\alpha}}^*$) can be partitioned into two blocks $\hat{\boldsymbol{\alpha}}_1$ and $\hat{\boldsymbol{\alpha}}_2$ (resp. $\hat{\boldsymbol{\alpha}}^*_1$ and $\hat{\boldsymbol{\alpha}}^*_2$) of dimensions $q \times q$ and $(p-q) \times q$:

$$\begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \hat{\boldsymbol{\alpha}}_2 \end{pmatrix} = \mathbf{H} \begin{pmatrix} \hat{\boldsymbol{\alpha}}^*_1 \\ \hat{\boldsymbol{\alpha}}^*_2 \end{pmatrix}$$

It remains to show that if at least one constraint of a column is different from zero, then the sign of this column is determined. Let $\hat{\boldsymbol{\alpha}}_{\cdot j}$ (resp. $\hat{\boldsymbol{\alpha}}^*_{\cdot j}$) be the j^{th} column of $\hat{\boldsymbol{\alpha}}$ (resp. $\hat{\boldsymbol{\alpha}}^*$) and let $\hat{\alpha}_{i'j'}$ be an element of the upper triangle of $\hat{\boldsymbol{\alpha}}$. Assume that it is different from zero, which means

$$\hat{\alpha}_{i'j'} = \hat{\alpha}^*_{i'j'} = a \neq 0.$$

Then, $\hat{\boldsymbol{\alpha}} = \mathbf{H}\hat{\boldsymbol{\alpha}}^*$ implies that $\hat{\alpha}_{i'j'} = h_{i'i'}\hat{\alpha}^*_{i'j'} = a$ and $h_{i'j'} = 1$. Hence, the sign of the j^{th} column of $\hat{\boldsymbol{\alpha}}$ is determined.

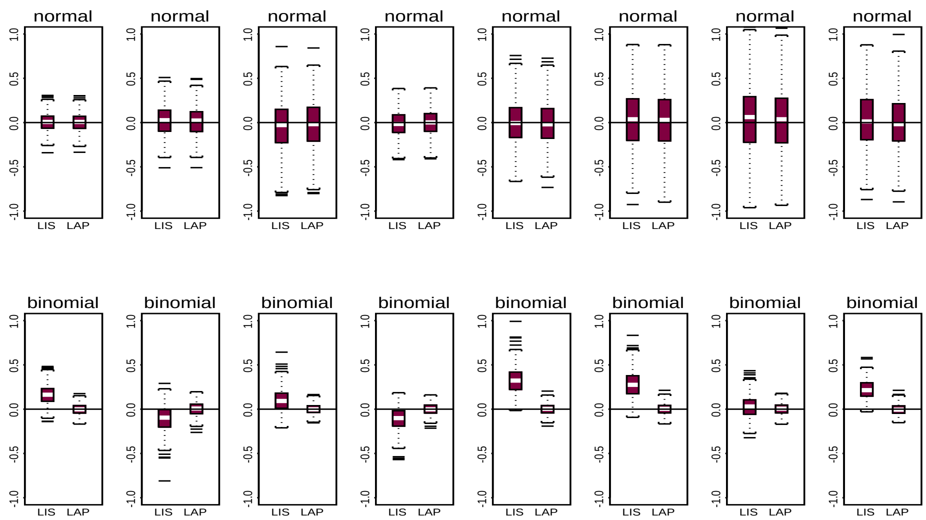

Figure 1: Estimation of $\boldsymbol{\alpha}_1$ for a model with four latent variables

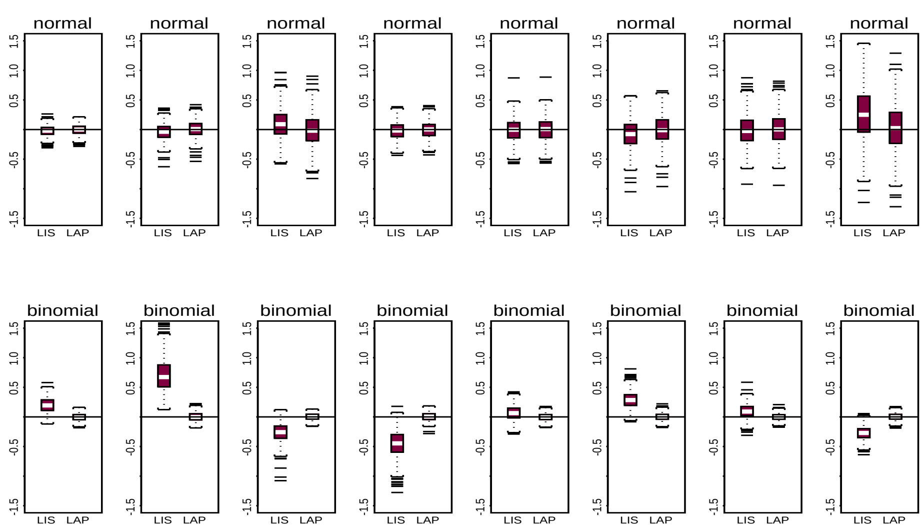

Figure 2: Estimation of $\boldsymbol{\alpha}_3$ for a model with four latent variables

References

Arminger, G. and Küsters, U. (1988) *Latent Trait Models.* New York: Plenum Press.

Bartholomew, D. J. (1984a) The foundations of factor analysis. *Biometrika*, **71**, 221–232.

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

Bartholomew, D. J. and Knott, M. (1999) *Latent Variable Models and Factor Analysis.* Edward Arnold Publishers Ltd.

Breslow, N. E. and Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.

Davison, A. C. (1992) Treatment effect heterogeneity in paired data. *Biometrika*, **79**, 463–474.

De Bruijn, N. G. (1981) *Asymptotic Methods in Analysis.* New York: Dover Publications, third edn.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (C/R: p22-37). *Journal of the Royal Statistical Society, Series B*, **39**, 1–22.

Dennis, J. E., J. and Schnabel, R. B. (1983) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* New York: Prentice-Hall.

Green, P. J. (1987) Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, **55**, 245–259.

Heritier, S. and Ronchetti, E. (1994) Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association*, **89**, 897–904.

Huber, P., Ronchetti, E. and Victoria-Feser, M.-P. (2003) Estimation of generalized latent trait models. *Cahiers du Département d'Econométrie 2003.05*, University of Geneva.

Huber, P. J. (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73–101.

— (1981) *Robust Statistics.* New York: Wiley.

Jöreskog, K. G. (1969) A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, **34**, 183–203.

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

Jöreskog, K. G. and Sörbom, D. (1993) *LISREL 8: Structural Equation Modeling with IMPLIS Command Language*. Hove London: Erlbaum.

Lee, Y. and Nelder, J. A. (1996) Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*, **58**, 619–678.

Lin, X. and Breslow, N. E. (1996) Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, **91**, 1007–1016.

Lord, F. M. and Novick, M. E. (1968) *Statistical Theories of Mental Test Scores*. Addison-Wesley Publishing Co.

Maasoumi, E. (1986) The measurement and decomposition of multi-dimensional inequality. *Econometrica*, **54**, 991–997.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. New York: Chapman & Hall, second edn.

McGilchrist, C. A. (1994) Estimation in generalized mixed models. *Journal of the Royal Statistical Society, Series B*, **56**, 61–69.

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

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2002) Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, **2**, 1–21.

— (2004) Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *manuscript*.

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

Sen, A. K. (1987) *On Ethics and Economics*. The Royer Lectures. Oxford: B. Blackwell.

Tierney, L. and Kadane, J. B. (1986) Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82–86.

Welsh, A. H. (1996) *Aspects of Statistical Inference*. New York: Wiley.

Yau, K. K. W. and McGilchrist, C. A. (1996) Simulation study of the GLMM method applied to the analysis of clustered survival data. *Journal of Statistical Computation and Simulation*, **55**, 189–200.

	Normal items			
	Income	Housing	Clothing	Leisure
	Constants			
LAMLE	9.134 (0.019)	7.218 (0.020)	5.803 (0.023)	6.138 (0.020)
LISREL	9.133	7.217	5.801	6.137
	Latent variable 1			
LAMLE	1.110 (0.013)	1.047 (0.020)	1.229 (0.019)	1.210 (0.017)
LISREL	1.106	1.051	1.226	1.211
	Latent variable 2			
LAMLE	0.069 (0.024)	-0.055 (0.030)	0.129 (0.034)	0.054 (0.031)
LISREL	0.134	0.006	0.193	0.118
	Standard deviations			
LAMLE	0.555 (0.005)	1.014 (0.005)	1.033 (0.008)	0.813 (0.007)
LISREL	0.556	1.013	1.033	0.811

	Binary items		
	Freezer	Dishwasher	Washing
	Latent variable 1		
LAMLE	0.222 (0.033)	0.463 (0.038)	0.175 (0.046)
LISREL	0.064	0.139	-0.046
	Latent variable 2		
LAMLE	0.764 (0.041)	0.895 (0.046)	0.553 (0.056)
LISREL	0.307	0.310	0.575

	Binary items		
	TV	Video	Car
	Latent variable 1		
LAMLE	-0.110 (0.077)	0.109 (0.040)	0.628 (0.055)
LISREL	-0.029	0.014	0.110
	Latent variable 2		
LAMLE	1.898 (0.122)	1.092 (0.063)	1.264 (0.057)
LISREL	0.236	0.392	0.308

Table 2: Parameter estimates for the consumption data with standard errors between parenthesis when available.