Estimation of Generalized Linear Latent Variable Models

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

No 2003.05

Cahiers du département d'économétrie
Faculté des sciences économiques et sociales
Université de Genève

Mai 2003

Département d'économétrie
Université de Genève, 40 Boulevard du Pont-d'Arve, CH -1211 Genève 4
http://www.unige.ch/ses/metri/

Estimation of Generalized Linear Latent Variable Models

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

May 2003

Abstract

Generalized Linear Latent Variable Models (GLLVM), as defined in Bartholomew and Knott (1999) allow to model relationships between manifest and latent variables when the manifest variables are of various type, such as continuous or discrete. They extend structural equation modelling techniques which are very powerful modelling tools in the social sciences. However, because of the complexity of the log-likelihood function of GLLVM due to the fact that the latent variables are not directly observed, usually an approximation such as numerical integration is used to carry out estimation and inference. This can limit in a drastic way 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. It is based on a Laplace approximation of the likelihood function and can be computed even for models with a large number of variables. It is shown that the new estimator can be viewed as a M-estimator leading to readily available asymptotic properties and correct inference. A simulation study in various settings shows its excellent finite sample properties, in particular when compared with a well established approach such as LISREL.

[*]Department of Econometrics, Faculty of Economic and Social Sciences, Blv. Pont d'Arve 40, CH-1211 Geneva, Switzerland

[¶]HEC, Faculty of Economic and Social Sciences and Faculty of Psychology and Education, 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. These concepts 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. Indeed, how are, for example, the welfare of people or the intelligence 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 been available for a long time; see e.g. Jöreskog (1969), Bartholomew (1984a) and Arminger and Küsters (1988). Factor analysis is one of them. A model is proposed to link manifest variables (supposed to be multivariate normal) with latent variables (or factors) and a likelihood analysis can be carried out. Since the work of Jöreskog (1969), a lot of 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 standalone softwares, such as LISREL; cf. Jöreskog (1990) and Jöreskog and Sörbom (1993).

Although LISREL is a package that incorporates methods dealing with a large variety of applied problems, it suffers from an important drawback in that 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 underlying indirect observations of multivariate normal variables. In other words, many applied problems are forced into the multivariate normal framework for which many statistical procedures have been developed.

In our opinion, it is essential that the manifest variables are treated as they are, i.e. binary, ordinal or continuous. The model that formalizes the relationship between the manifest and the latent variables should take the type of data at hand into account. These types of 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 mixture of manifest variables. They proposed a generalized linear latent variable model (GLLVM) (see Bartholomew and Knott, 1999) that allows to link latent variables to manifest variables of different type (see section 2.2).

However, the statistical analysis of GLLVM presents a difficulty. Since the latent variables are not observed, they must be integrated out from the likelihood function. Moustaki (1996) and Moustaki and Knott (2000) propose to use Gauss-Hermite Quadrature as a numerical approximation method. As it will be shown later, this implies that the possible number of estimable latent variables included in the model is restricted at the moment only to two.

In this paper, we instead propose the Laplace approximation for the likelihood function. This technique has three important advantages with respect to the quadrature method. First of all, it allows to derive the statistical properties of the estimator and to carry out valid inference. Second, it allows to estimate more complex models, in particular models with more than two latent variables, as well as models with correlated latent variables. Third, it allows direct estimation of individual scores on the latent variables space (see section 3.3).

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, as well as the GLLVM introduced by Bartholomew and Moustaki. In section 3, we propose a new estimator for the GLLVM based on the Laplace approximation of the likelihood function and investigate its statistical properties. The explicit formulas 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.

Finally, we compare our estimator with the ones provided by LISREL and the GLLVM with the Gauss-Hermite Quadrature in section 5. This clearly reveals that the new estimator has better performance in terms of bias and variance. As a conclusion, an overview of possible extensions is given in section 6.

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 $Y|\mathbf{z}$ be an underlying normal variable with mean $\boldsymbol{\alpha}^T\mathbf{z}$ and unit variance. Given $\mathbf{z}$, a link is then established between $X|\mathbf{z}$ and $Y|\mathbf{z}$ in that it is assumed that $X|\mathbf{z}$ takes the value 1 if $Y|\mathbf{z}$ is positive and 0 otherwise. Then, the expected value of $X|\mathbf{z}$ is

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

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

$$\text{probit}(E(X|\mathbf{z})) = \boldsymbol{\alpha}^T\mathbf{z}.$$

In practice, like in LISREL, the model parameters are estimated in three steps (see for instance 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. 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 ($|\Phi(x) - \Psi(1.7x)| < 0.01$ $\forall x$, where Ψ is the logistic distribution function, see e.g. 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).

2.2 Generalized Linear Latent Variable Model (GLLVM)

In this section, we present the GLLVM starting from the framework of generalized linear models (GLM); cf. McCullagh and Nelder (1989). The aim 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 $\boldsymbol{\alpha}_j = [\alpha_{j0}, \ldots, \alpha_{jq}]^T$ 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 $\prod_{j=1}^{p} g_j(x^{(j)}|\mathbf{z})$. Without loss of generality, it is also assumed that the distribution of the latent variables is the standard normal and that latent variables are independent. The last assumption can be relaxed (see section 3). Thus, the density $h(\mathbf{z}_{(2)})$ of $\mathbf{z}_{(2)}$ is the multivariate normal with mean $\boldsymbol{\mu} = 0$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_q$, the identity matrix of dimension q. The joint distribution of the manifest and latent variables is given by

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

However, since the latent variables $\mathbf{z}_{(2)}$ are not observed, their realizations are treated as missing. Therefore, they are integrated out so that one actually considers the marginal distribution $f_{\boldsymbol{\alpha},\boldsymbol{\phi}}(\mathbf{x})$, $\mathbf{x} = [x^{(1)}, \ldots, x^{(p)}]$, of the manifest variables given by

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

Note that $g_j(x^{(j)}|\mathbf{z})$ may be either normal or binomial according to j (or, indeed, another distribution from the exponential family). The aim is to obtain estimators for the parameters $\boldsymbol{\alpha}_j$ and ϕ_j, with $j = 1 \ldots p$. Once these estimators are known, any response pattern can be linked to values of the latent variables.

Note also that (3) formulates the general approach used with missing values (see e.g. Dempster, Laird, and Rubin, 1977). However, an explicit expression for (3) avoiding the integration doesn't exist, so that a numerical approximation is needed. Then, the EM algorithm can be used to find the (approximated) MLE of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$, as it is for example pointed out in Sammel, Ryan, and Legler (1997).

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

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

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

Equation (4) contains a multidimensional integral which cannot be computed explicitly, except when all the distributions $g_j(x^{(j)}|\mathbf{z})$ are normal. Consequently, an approximation of this integral is needed. Then depending on the chosen approximation, the estimators will be different in that their performance in terms of bias and variance is different.

3 Estimators based on the Laplace approximation (LAMLE)

The Gauss-Hermite Quadrature (GHQ) approximation to the integral in (4) proposed by Moustaki (1996) is easy to implement but suffers from 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 with GHQ. Secondly, making correct inference on the resulting estimators seems to be very difficult. Finally, the resulting estimator appears to be biased; cf. 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 (more latent variables imposing 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 variables space (see section 3.3). Finally, in our simulations, we found that it leads to unbiased estimators; cf. 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

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

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) (cf. De Bruijn (1981) or Tierney and Kadane (1986, pp. 82-86)), we obtain

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

where

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

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

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

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

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

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

3.2 LAMLE

The Laplace approximation allows to eliminate the integral from the marginal distribution of $\mathbf{x}_i$. From (6), (7), (8), and (9), we obtain the approximate log-likelihood function

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

The new estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ based on the Laplace approximation (LAMLE) 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

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

Hence, (12) and (13) provide a set of $p(q+2)$ estimating equations defining the estimators for the model parameters. In addition, (10) is required for the computation of all $n\cdot q$ 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 computations for these cases are tedious but straightforward. The LAMLE can be computed in principle for any mixture of distributions from the exponential family by using (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 LAMLE

A way to interpret the estimators derived in section 3.2 is to consider the $\hat{\mathbf{z}}_{i(2)}$ as "parameters" in (2). Then the "likelihood" would be

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

which differs from (11) by the additive factor $-\frac{1}{2}\sum_{i=1}^{n}\log\det(\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{z}_i))$. Taking the derivative of l^* with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ doesn't lead to the same expressions for the score function and hence, the 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 on the latent variables space.

3.4 Statistical properties of the LAMLE

Let $\hat{\boldsymbol{\theta}}_L$ be the vector containing all the LAMLE of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ for a GLLVM. $\hat{\boldsymbol{\theta}}_L$ 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 implicitely 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

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

These conditions must be checked in each particular model.

Moreover, the function $\tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x})$ in (11) plays the role of a pseudo-likelihood function and it 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 to carry out inference and variable selection in GLLVM.

4 Constraints and correlated latent variables

The estimating equations which define the LAMLE, or the MLE, 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 LAMLE

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

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

in which $\nu(\cdot)$ 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, $\mathbf{z}^*_{(2)} = \mathbf{H}\mathbf{z}_{(2)}$ has the same property. 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 same problem encountered in factor analysis, for example. If a unique solution is required, it is necessary to impose constraints on the parameters$\boldsymbol{\alpha}$.

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 thej^{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 LAMLE of a GLLVM with correlated latent variables

The flexible form of the Laplace approximation allows 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

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

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

$$Q(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{z}, \mathbf{x}_i) = \frac{1}{p}\left[\sum_{j=1}^{p}\left[\frac{x_i^{(j)}\boldsymbol{\alpha}_j^T\mathbf{z} - b_j(\boldsymbol{\alpha}_j^T\mathbf{z})}{\phi_j} + c_j(x_i^{(j)}, \phi_j)\right] - \frac{\mathbf{z}_{(2)}^T\boldsymbol{\Sigma}^{-1}\mathbf{z}_{(2)}}{2} - \frac{q}{2}\log(2\pi)\right] \tag{15}$$

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

$$\hat{\mathbf{z}}_{i(2)} := \hat{\mathbf{z}}_{i(2)}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{x}_i, \boldsymbol{\Sigma}) = \boldsymbol{\Sigma}\sum_{j=1}^{p}\frac{1}{\phi_j}\left(x_i^{(j)}\boldsymbol{\alpha}_{j(2)} - \frac{\partial b_j(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{\partial \mathbf{z}_{i(2)}}\right). \tag{16}$$

The LAMLE estimating equations with correlated latent variables are the modified (12) and (13) using (15) and, in addition, the $q(q-1)/2$ equations for the elements σ_{kl} of $\boldsymbol{\Sigma}$:

$$\begin{aligned}\frac{\partial \tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x})}{\partial \sigma_{kl}} = \sum_{i=1}^{n}\Bigg[&-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i, \boldsymbol{\Sigma})^{-1}\frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i, \boldsymbol{\Sigma})}{\partial \sigma_{kl}}\right)\\ &-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{kl}}\right) + \frac{1}{2}\hat{\mathbf{z}}_{i(2)}^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{kl}}\boldsymbol{\Sigma}^{-1}\hat{\mathbf{z}}_{i(2)}\Bigg] = 0.\end{aligned} \tag{17}$$

5 Numerical application

In this section, we compare the LAMLE (with uncorrelated latent variables) with the MLE using the GHQ approximation and the LISREL estimators that we take as the benchmarks. We consider a model containing one, two, and four latent variables. The code to compute the MLE using the GHQ approximation was kindly provided by I. Moustaki. 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.

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 with LISREL estimators because they can be interpreted as GLM 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 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.

5.2 Estimation

With one latent variable, it is possible to use up to 48 quadrature points with GHQ. Thus, we are now comparing a GHQ approximation with 48 quadrature points with the Laplace approximation and the LISREL method. We created 500 samples of size 200 with 5 manifest variables, 3 of them are normal and 2 are binomial. The values of the parameters are presented in Table 1. Note that with other parameters values, we found similar results.

	Normal			Binomial	
$\boldsymbol{\alpha}_0$	2.4	3.5	2.8	2.5	1.5
$\boldsymbol{\alpha}_1$	3.3	3.6	3.5	0.7	0.5
$\boldsymbol{\phi}$	1.0	12.0	3.0		

Table 1: Parameters for a model with one latent variable

Two sets of boxplots are presented in Figures 1 and 2: the first one for the estimators of $\boldsymbol{\alpha}_1$ and the second one for the scale parameters $\boldsymbol{\phi}$. A similar plot was obtained for $\boldsymbol{\alpha}_0$ (not shown here). Each set represents three boxplots. The left ones correspond to the LISREL estimators, the center ones to the MLE with a GHQ approximation with 48 quadrature points and the right ones to the LAMLE. Estimates have been centered by substracting the true parameters values. Results are discussed in section 5.3.

With two latent variables, we use 8 and 16 quadrature points (to go beyond 16 points would be too computer intensive). Again, 500 samples of size 200 were generated. They are built with 5 normal and 5 binomial manifest variables. The parameter α_{21} is set to zero (i.e. not estimated).

	Normal items					Binomial items				
$\boldsymbol{\alpha}_0$	5.0	-2.0	3.0	0.0	-8.0	2.0	-1.5	-1.2	-0.5	0.2
$\boldsymbol{\alpha}_1$	4.0	2.0	-6.0	1.0	-3.0	0.1	0.0	-1.5	-0.8	-0.3
$\boldsymbol{\alpha}_2$	—	6.0	4.0	8.0	-2.0	-2.3	0.5	1.4	0.1	0.0
$\boldsymbol{\phi}$	1.0	1.5	2.0	3.0	0.5					

Table 2: Parameters for a model with two latent variables

In Figures 3 and 4, we present the distributions of the estimators using four boxplots. Each corresponds respectively (from left to right) to the LISREL estimators, the MLE using the GHQ approximation with 8 and 16 quadrature points and the LAMLE. We discuss the results in section 5.3. It should be stressed that other sets of parameters produced similar results.

Finally, models with four latent variables can only be estimated by the MLE using the Laplace approximation and the LISREL. 500 samples of size 400 were simulated. They contain 8 normal and 8 binomial responses. The parameters are given in Table 3. The parameters α_{21}, α_{31}, α_{22}, α_{41}, α_{42}, and α_{43} are set to zero. Figures 5 and 6 present the

	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
$\boldsymbol{\alpha}_2$	—	1.0	-3.0	-2.0	0.0	3.0	4.0	5.0
$\boldsymbol{\alpha}_3$	—	—	3.0	0.0	-1.0	2.0	4.0	-9.0
$\boldsymbol{\alpha}_4$	—	—	—	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 3: Parameters for a model with four latent variables

results, with paired boxplots. The left ones correspond to the LISREL estimators and the right ones to the LAMLE.

5.3 Discussion of the results

Models with a single latent variable are rather simple and the GHQ approximation with 48 quadrature points is expected to give good results. Figures 1 and 2 show no differences between the GHQ approximation and the LAMLE. Even, in this model, with only 5 manifest variables, the LAMLE shows a performance of the same quality as the GHQ approximation. Variances of the estimators are very close and no bias appears. The LISREL estimator for the first binomial loading looks however biased. In conclusion, LAMLE is as good as the MLE with the GHQ approximation for very low dimensional models for GLLVM and the LISREL already shows problem with binomial manifest variables.

Insert Figures 1 and 2 here

In models containing two latent variables, the quality of GHQ is expected to deteriorate because the implementation allows only for 16 quadrature points at most. On the other hand, the LAMLE behavior should not change as p grows to 10. Large biases appear with GHQ approximations for the parameters (see Figures 3 and 4). For instance, ϕ_{33} reveals a bias that is so large, that every sample leads to an estimate always above the true value. On the other hand, the LAMLE remain unbiased.

Insert Figures 3 and 4 here

A possible explanation of this difference is as follows. The GHQ approximation is based upon the integration on pre-specified quadrature points: these points are placed on a grid which is fixed and does not depend on the form of the log-likelihood. With 16 and 8 quadrature points, this grid becomes coarser. Hence, it can happen that the peak of the log-likelihood lies right in a hole; see Figure 7. In such a case, most of the information contained in 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.

———————————————————————

Insert Figure 7 here

———————————————————————

The LISREL estimators are unbiased for normal manifest variables but show important biases for some of the binomial manifest variables (see α_{18} and α_{26} for example).

With four latent variables, no comparisons between GHQ and the LAMLE are possible. Actually, the most important fact is that the LAMLE is easily computable on dimensions that were untractable for GLLVM before. The results of the estimators of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_3$ are presented in Figures 5 and 6. As it was already the case for one and two latent variables, LAMLE are 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's values.

———————————————————————

Insert Figures 5 and 6 here

———————————————————————

6 Conclusion

The aim of this paper was to propose a general method for estimating the parameters of a GLLVM even in high dimensional models. The likelihood function of a GLLVM contains integrals that need to be approximated. Moustaki (1996) proposed a GHQ approximation which, although simple, can lead to biased estimates. Moreover, with such an approximation, the GLLVM model is limited to two latent variables. We proposed instead to use a Laplace approximation for the likelihood function which allows to estimate models with many latent variables. We showed that the LISREL approach can lead to highly biased estimators for the loadings of binomial manifest variables whereas the LAMLE remain unbiased. The estimators based on the Laplace approximation can be interpreted as M-estimators, and as a consequence, inference is readily available. Moreover, the estimation procedure provides automatically individual scores on the latent variables space. All the procedures presented in this paper are implemented in a standalone software which is available from the authors upon request. Open research directions include the development of variable selection procedures as in GLM based on the LAMLE.

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

A.1 Binomial manifest variables

Let $X|\mathbf{Z}$, 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 given by

$$\begin{aligned} b(\boldsymbol{\alpha}^T\mathbf{z}) &= m\text{log}(1 + \exp(\boldsymbol{\alpha}^T\mathbf{z})) && \text{(18a)} \\ c(x, \phi) &= c(x) = \log\begin{pmatrix} m \\ x \end{pmatrix}, && \text{(18b)} \end{aligned}$$

and

$$g(x|\mathbf{z}) = \begin{pmatrix} m \\ x \end{pmatrix} \pi(\mathbf{z})^x (1 - \pi(\mathbf{z}))^{(m-x)}. \tag{19}$$

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

$$\begin{aligned} \tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x}) = \sum_{i=1}^{n} \Bigg(&-\frac{1}{2}\text{log det}\big(\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i)\big) \\ &+ \sum_{j=1}^{p} \left[x_i^{(j)} \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i - m\text{log}(1 + \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)) + \log\begin{pmatrix} m \\ x_i^{(j)} \end{pmatrix} \right] - \frac{\hat{\mathbf{z}}_{i(2)}^T \hat{\mathbf{z}}_{i(2)}}{2} \Bigg), \end{aligned} \tag{20}$$

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

$$\hat{\mathbf{z}}_{i(2)} = \sum_{j=1}^{p} \left(x_i^{(j)} - m \frac{\exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)}{1 + \exp(\boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_i)} \right) \boldsymbol{\alpha}_{j(2)}. \tag{21}$$

To compute the score functions, we first need

$$\mathrm{tr}\left(\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i)^{-1}\frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i)}{\partial \alpha_{kl}}\right) = \mathrm{tr}\Bigg(\left(\sum_{j=1}^{p} m\beta_{ji}\boldsymbol{\alpha}_{j(2)}\boldsymbol{\alpha}_{j(2)}^T + \mathbf{I}_q\right)^{-1}$$
$$\left(\sum_{j=1}^{p} m\beta_{ji}\left[\frac{1-\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}\frac{\partial \boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i}{\partial \alpha_{kl}}\boldsymbol{\alpha}_{j(2)}\boldsymbol{\alpha}_{j(2)}^T\right] + (1-\delta_{l0})(\mathbf{e}_l \otimes \boldsymbol{\alpha}_k^T + \mathbf{e}_l^T \otimes \boldsymbol{\alpha}_k)\right)\Bigg), \tag{22}$$

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,

$$\frac{\partial b_j(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{\partial \boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i} = m\frac{\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}. \tag{23}$$

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

$$\frac{\partial \hat{\mathbf{z}}_{i(2)}}{\partial \alpha_{k0}} = -m\beta_{ki}\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i)^{-1}\boldsymbol{\alpha}_{k(2)} \tag{24a}$$
$$\frac{\partial \hat{\mathbf{z}}_{i(2)}}{\partial \boldsymbol{\alpha}_{k(2)}} = \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i)^{-1}\left(-m\beta_{ki}\boldsymbol{\alpha}_{k(2)}\hat{\mathbf{z}}_{i(2)}^T + \left(x_i^{(k)} - m\frac{\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}\right)\mathbf{I}_q\right). \tag{24b}$$

The LAMLE of a model with binomial manifest variables is completely defined by the pseudo log-likelihood (20) and its score functions (12) whose components are given by (21), (22), (23), and (24).

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:

$$\tilde{l}(\boldsymbol{\alpha}, \boldsymbol{\phi}|\mathbf{x}) = \sum_{i=1}^{n}\Bigg(-\frac{1}{2}\log\,\det\big(\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)\big)$$
$$+\sum_{j=1}^{p_1}\left[\frac{\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i}{\phi_j}\left(x_i^{(j)} - \frac{\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i}{2}\right) - \frac{1}{2}\left(\frac{\left(x_i^{(j)}\right)^2}{\phi_j} + \log(2\pi\phi_j)\right)\right]$$
$$+\sum_{j=p_1+1}^{p}\left[x_i^{(j)}\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i - m\log(1+\exp(\boldsymbol{\alpha}_j^T\cdot\hat{\mathbf{z}}_i)) + \log\begin{pmatrix} m \\ x_i^{(j)}\end{pmatrix}\right] - \frac{\hat{\mathbf{z}}_{i(2)}^T\hat{\mathbf{z}}_{i(2)}}{2}\Bigg), \qquad (25)$$

where

$$\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i) = \sum_{j=1}^{p_1}\frac{\boldsymbol{\alpha}_{j(2)}\boldsymbol{\alpha}_{j(2)}^T}{\phi_j} + \sum_{j=p_1+1}^{p} m\beta_{jk}\boldsymbol{\alpha}_{j(2)}\boldsymbol{\alpha}_{j(2)}^T + \mathbf{I}_q = \boldsymbol{\Gamma_1}(\boldsymbol{\alpha}, \boldsymbol{\phi}) + \boldsymbol{\Gamma_2}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i) + \mathbf{I}_q.$$

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

$$\hat{\mathbf{z}}_{i(2)} = \sum_{j=1}^{p_1}\frac{1}{\phi_j}(x_i^{(j)} - \boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)\boldsymbol{\alpha}_{j(2)} + \sum_{j=p_1+1}^{p}\left(x_i^{(j)} - m\frac{\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}\right)\boldsymbol{\alpha}_{j(2)}. \qquad (26)$$

We differentiate (25) 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

$$\frac{\partial\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)}{\partial\alpha_{kl}} = (1-\delta_{l0})\big(\mathbf{e}_l\otimes\boldsymbol{\alpha}_i^T + \mathbf{e}_l^T\otimes\boldsymbol{\alpha}_i\big)\big(\frac{1}{\phi_k}D_1 + m\beta_{ki}D_2\big)$$
$$+\sum_{j=p_1+1}^{p} m\beta_{ji}\left(\frac{1-\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}\frac{\partial\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i}{\partial\alpha_{kl}}\boldsymbol{\alpha}_{j(2)}\boldsymbol{\alpha}_{j(2)}^T\right), \qquad (27a)$$

where

$$D_1 = \begin{cases} 1 & : \quad 1\le i\le p_1 \\ 0 & : \quad p_1 < i \le p\end{cases} \qquad \text{and} \qquad D_2 = \begin{cases} 0 & : \quad 1\le i\le p_1 \\ 1 & : \quad p_1 < i \le p\end{cases},$$

and

$$\frac{\partial\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)}{\partial\phi_k} = -\frac{1}{\phi_k^2}\boldsymbol{\alpha}_{k(2)}\boldsymbol{\alpha}_{k(2)}^T + \sum_{j=p_1+1}^{p} m\beta_{ji}\frac{1-\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}\frac{\partial\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i}{\partial\phi_k}\boldsymbol{\alpha}_{j(2)}\boldsymbol{\alpha}_{j(2)}^T. \qquad (27b)$$

Moreover,

$$\frac{\partial\hat{\mathbf{z}}_{i(2)}}{\partial\alpha_{k0}} = \begin{cases} -\frac{1}{\phi_k}\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\phi})^{-1}\boldsymbol{\alpha}_{k(2)}, & \text{if } 1\le i\le p_1 \\ -m\beta_{ki}\boldsymbol{\Gamma}(\boldsymbol{\alpha}, \hat{\mathbf{z}}_i)^{-1}\boldsymbol{\alpha}_{k(2)}, & \text{otherwise}\end{cases} \qquad (28a)$$

$$\frac{\partial \hat{\mathbf{z}}_{i(2)}}{\partial \alpha_{kl}} = \begin{cases} \frac{1}{\phi_k}\boldsymbol{\Gamma}(\boldsymbol{\alpha},\boldsymbol{\phi})^{-1}\Big(-\boldsymbol{\alpha}_{k(2)}\hat{\mathbf{z}}_{i(2)}^T + (x_i^{(k)} - \boldsymbol{\alpha}_k^T\hat{\mathbf{z}}_i)\mathbf{I}_q\Big), & \text{if } 1 \le i \le p_1 \\ \boldsymbol{\Gamma}(\boldsymbol{\alpha},\hat{\mathbf{z}}_i)^{-1}\Big(-m\beta_{ki}\boldsymbol{\alpha}_{k(2)}\hat{\mathbf{z}}_{i(2)}^T + \big(x_i^{(k)} - m\frac{\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}{1+\exp(\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_i)}\big)\mathbf{I}_q\Big), & \text{otherwise} \end{cases} \tag{28b}$$

$$\frac{\partial \hat{\mathbf{z}}_{i(2)}}{\partial \phi_k} = -\boldsymbol{\Gamma}(\boldsymbol{\alpha},\boldsymbol{\phi})^{-1}\left(\frac{1}{\phi_k^2}(x_i^{(k)} - \boldsymbol{\alpha}_i^T\hat{\mathbf{z}}_i)\boldsymbol{\alpha}_{k(2)}\right), \quad \text{if } 1 \le i \le p_1. \tag{28c}$$

Thus, the pseudo log-likelihood (25) is maximized when the score functions given by (12) and (13) are set to zero, where expressions (21), (27) and (28) are used.

Appendix B: Proof of Proposition 1

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

Let $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_{ij})_{1\le i,j\le q}$ and $\hat{\boldsymbol{\alpha}}^* = (\hat{\alpha}^*_{ij})_{1\le i,j\le q}$ be two square matrices of dimension $q \times q$ and $\mathbf{H} = (h_{ij})_{1\le i,j\le q}$ 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 $1 \le i, j \le q$ 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.

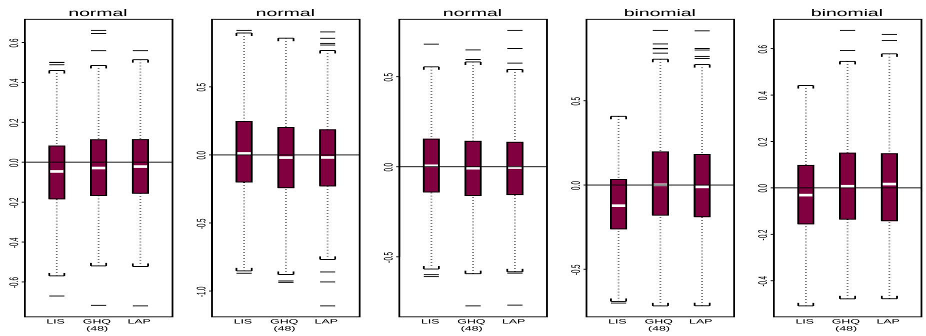

Figure 1: Estimation of $\boldsymbol{\alpha}_1$ for a model with a single latent variable

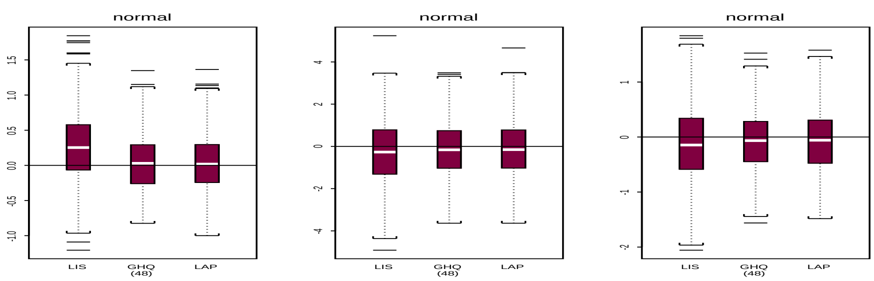

Figure 2: Estimation of the scale parameters $\boldsymbol{\phi}$ for a model with a single latent variable

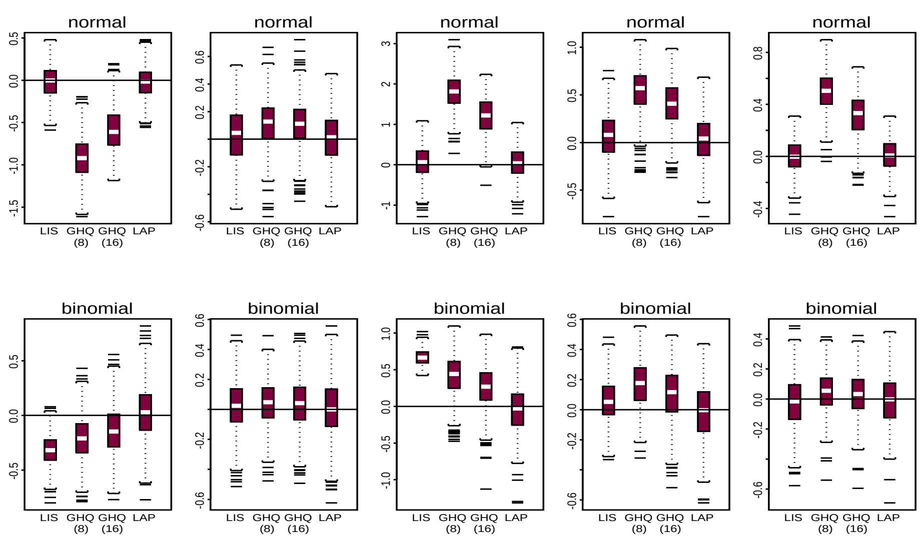

Figure 3: Estimation of $\boldsymbol{\alpha}_1$ for a model with two latent variables

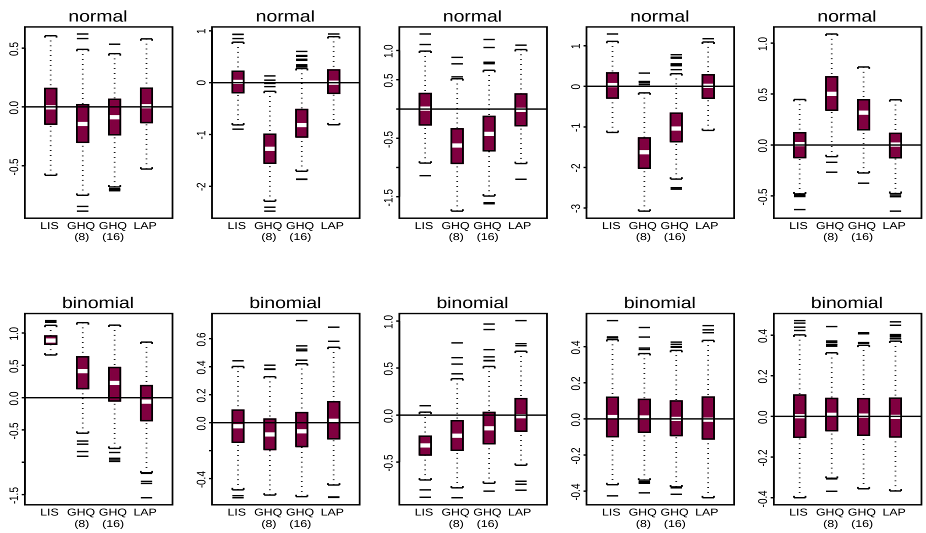

Figure 4: Estimation of $\boldsymbol{\alpha}_2$ for a model with two latent variables

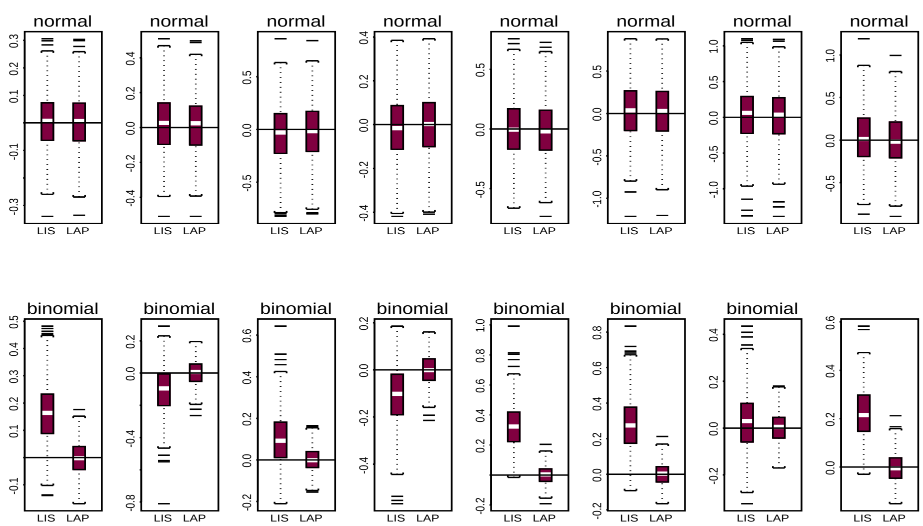

Figure 5: Estimation of $\boldsymbol{\alpha}_1$ for a model with four latent variables

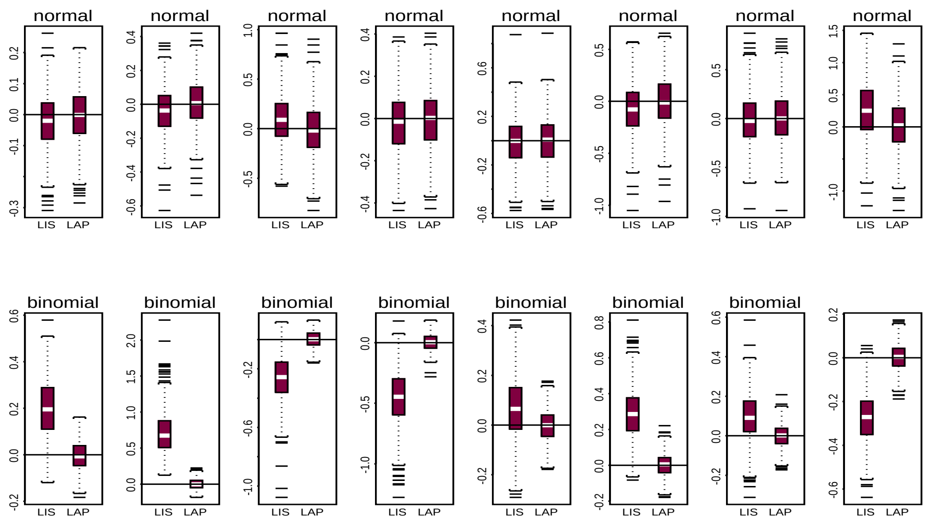

Figure 6: Estimation of $\boldsymbol{\alpha}_3$ for a model with four latent variables

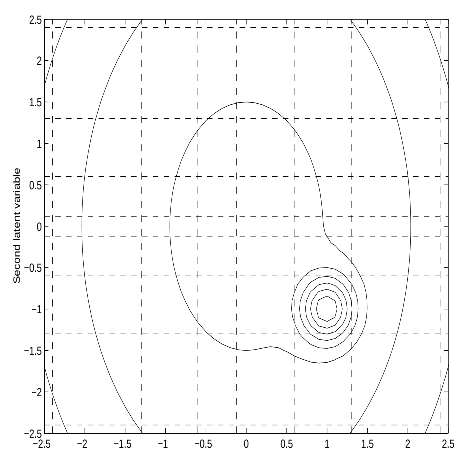

Figure 7: Log-likelihood that would be badly approximated by GHQ (contour view). The grid's knots are the quadrature points

References

Arminger, G. and Küsters, U. (1988), *Latent Trait Model*, 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.

De Bruijn, N. G. (1981), *Asymptotic Methods in Analysis*, New York: Dover Publications, 3rd ed.

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.

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. J. (1964), "Robust Estimation of a Location Parameter," *The 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.

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

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, New York: Chapman & Hall, 2nd ed.

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.

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.