Parameter constraints in generalized linear latent variable models

R. Tsonaka\textsuperscript{a}, I. Moustaki\textsuperscript{b,*}

\textsuperscript{a}Biostatistical Centre, Catholic University of Leuven, Belgium
\textsuperscript{b}Department of Statistics, Athens University of Economics and Business, 76 Patission street, 104 34 Athens, Greece

Received 25 March 2005; received in revised form 8 April 2006; accepted 10 April 2006
Available online 11 May 2006

Abstract
Parameter constraints in generalized linear latent variable models are discussed. Both linear equality and inequality constraints are considered. Maximum likelihood estimators for the parameters of the constrained model and corrected standard errors are derived. A significant reduction in the dimension of the optimization problem is achieved with the proposed methodology for fitting models subject to linear equality constraints.

Keywords: Generalized linear latent variable models; Linear equality and inequality constraints; Lagrange multipliers; Adaptive barrier method

1. Introduction

Latent variable models are widely used in social science research where variables of major interest such as ability, attitudes, behavior, cannot be directly measured. For example, in educational testing the performance of students on a number of tests is used as an indicator of ability or in economics variables such as income, expenditures, ownership of car, summer house, etc., are used to measure wealth. In some cases, a construct can be represented by a single latent variable but often it is multidimensional and therefore more than one latent variable is needed.

In latent variable modelling there is often the need to impose parameter constraints. More specifically, model parameters such as intercepts and factor loadings are constrained to be equal to or greater/less than a fixed value or other parameters according to some linear or non-linear function. Constraints are mainly required in confirmatory analysis where factor loadings follow a pre-specified pattern. In addition, fixed value constraints provide one way of identifying the scale of a latent variable (see Bollen, 1989, p. 183). Equality constraints among parameters in different groups are used for testing measurement invariance. Finally, constraints need to be imposed for making a model identified. Lack of identification implies that the model contains insufficient information for the purpose of attaining a determinate solution.

Fitting models under parameter constraints has been studied in structural equation modelling where limited information estimation methods are mainly used (Jöreskog, 1971; Bentler and Weeks, 1980; McDonald, 1980; Lee, 1980; Lee and Tsui, 1982; Bentler and Lee, 1983; Rindskopf, 1983, 1984; Jamshidian and Bentler, 1993). Constraints have also

\* Corresponding author.
E-mail addresses: spyridoula.tsonaka@med.kuleuven.be (R. Tsonaka), moustaki@aueb.gr (I. Moustaki).
been imposed in latent class models where full maximum likelihood estimation is used (Goodman, 1947a,b; Wright and Stone, 1979; Fischer, 1983; Formann, 1985; Clogg and Goodman, 1985; Haberman, 1988; van de Pol and Langeheine, 1990; Mooijaart and Heijden, 1992; Hooijtink, 1998). The software MULTILOG (Thissen et al., 2003) incorporates fixed value and equality constraints for the one, two and three-parameter logistic models.

We distinguish two types of linear constraints, namely equality and inequality constraints that are formulated, respectively, as

\[ V\mathbf{z} - b = 0 \] (1)

and

\[ U\mathbf{z} - d \geq 0, \] (2)

where \( V \) and \( U \) are known constraint matrices of order \((b \times v)\) and \((d \times v)\), respectively, \( \mathbf{z} \) is the parameter vector of length \((v \times 1)\) and \( b \) and \( d \) are known fixed-value vectors of length \((b \times 1)\) and \((d \times 1)\), respectively. According to (1), linear equality constraints include:

1. Fixed value constraints, that is one or more parameters are set equal to a fixed value (including the value zero).
2. Positive constraints, that is a parameter is set equal to other model parameter(s) (i.e., \( z_1 = z_2 \) or equivalently \( z_1 - z_2 = 0 \)).
3. Negative constraints, that is a parameter is set equal to the negative of other parameter(s) (i.e., \( z_1 = -z_2 \) or equivalently \( z_1 + z_2 = 0 \)).
4. Proportionality constraints, that is the ratio of two parameters is set equal to a fixed value (i.e., \( z_1/z_2 = c_1 \) or equivalently \( z_1 - c_1 z_2 = 0 \)).
5. Additive constraints, that is the sum of two parameters (or a linear function of them) is set equal to a fixed value (i.e., \( z_1 + c_1 z_2 = c_3 \)).

The linear inequality constraints are defined similarly to the linear equality constraints, i.e., we distinguish between fixed value, positive, negative, proportionality and additive inequality constraints.

In this paper, we propose a method for imposing linear equality parameter constraints in generalized linear latent variable models (GLLVM) that considerably reduces the dimension of the estimation. We also investigate more traditional ways of imposing constraints such as the Lagrange multipliers method (Bertsekas, 1982) and apply the adaptive barrier method (Lange, 1994) for handling inequality constraints.

The paper is organized as follows: Section 2 gives the theoretical framework of the GLLVM; Section 3 discusses maximum likelihood estimation under linear equality constraints and proposes a computationally efficient algorithm; Section 4 discusses maximum likelihood estimation under linear inequality constraints; Section 5 gives the formulas for estimating standard errors of the constrained estimated parameters. In Section 7, we evaluate the performance of the proposed methodology and compare it with the Lagrange multipliers method through a simulation study; Section 8 illustrates the methodology using two examples and Section 9 outlines the results of this work. Finally, in the Appendix we give the details of the computational framework given in Section 3.2.

2. Generalized linear latent variable models

We adopt here the GLLVM framework given in Moustaki and Knott (2000). That model framework includes binary, nominal and metric observed variables. More specifically, a GLLVM consists of three components:

1. The random component in which each of the \( p \) random \( \mathbf{x}^T = (x_1, \ldots, x_p) \) response variables given the latent variables \( \mathbf{z}^T = (z_1, \ldots, z_q) \) has a distribution from the exponential family.
2. The systematic component in which the latent variables produce a linear predictor \( \eta_i \) corresponding to each \( x_i \):

\[ \eta_i = \alpha_0 + \sum_{j=1}^{q} \alpha_{ij} z_j, \quad i = 1, \ldots, p, \quad j = 1, \ldots, q. \] (3)
The linear predictor for nominal variables with \( C_i \) response categories is written as

\[
\eta_{i(c)} = \eta_{0(c)} + \sum_{j=1}^{q} \alpha_{ij(c)} z_j, \quad i = 1, \ldots, p, \quad j = 1, \ldots, q, \quad c = 1, \ldots, C_i.
\]

(3) The links between the systematic component and the conditional means of the random component distributions

\[
\eta_i = \mu_i(z) \quad \text{with} \quad \mu_i(z) = E\left(x_i | z\right),
\]

where \( \mu_i(.) \) is called the link function which can be any monotonic differential function and may be different for different manifest variables \( x_i, i = 1, \ldots, p \). The results presented in this paper can be applied to any distribution of the exponential family and to any link function.

The conditional distributions of the manifest variables \( x_i \) given the latent variables, denoted by \( g_i(x_i | z) \) (\( i = 1, 2, \ldots, p \)) (the subscript \( i \) on \( g \) implies that items are allowed to have different distributions), are taken from the exponential family of distributions:

\[
g_i(x_i | z) = \exp\left\{ \frac{x_i \theta_i(z) - b_i(\theta_i(z))}{\phi_i} + c_i(x_i, \phi_i) \right\}.
\]

All the distributions discussed in this paper have canonical link functions with \( \theta_i = \eta_i \). For presentation clarity \( \theta_i(z) \) and \( b_i(\theta_i(z)) \) will be written as \( \theta_i \) and \( b_i(\theta_i) \), respectively. The \( \theta_i(\theta) \) and \( c_i(x, \phi) \) are specific functions taking a different form depending on the distribution of the response variable \( x_i \) and \( \phi \) is the scale parameter. For nominal observed variables, a subscript \( c \) is added on the functions \( \theta_i \) and \( b_i(\theta_i) \) that denotes the response category. Details of the GLLVM components for different types of observed variables can be found in Moustaki and Knott (2000).

Under the assumption of conditional independence (responses to the \( p \) items are independent conditional on the latent variables) the joint distribution of the manifest variables is

\[
f(x) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(z) \prod_{i=1}^{p} g_i(x_i | z) \, dz,
\]

where \( h(z) \) denotes the joint distribution of the latent variables. Latent variables are assumed to be independent and identically distributed with standard normal distributions.

For a random sample of size \( n \) the log-likelihood is

\[
\ell(\omega) = \sum_{m=1}^{n} \log f(x_m) = \sum_{m=1}^{n} \log \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(z) \prod_{i=1}^{p} \exp\left\{ \frac{x_{im} \theta_i - b_i(\theta_i)}{\phi_i} + c_i(\phi_i, x_{im}) \right\} \, dz,
\]

where \( \omega^T = (\theta, \phi)^T \) denotes the full parameter vector. Since constraints are imposed on the canonical parameter \( \theta \), the profile log-likelihood \( \ell(z) = \ell(z|\phi) \) is used instead. Besides, in the following sections we will discuss minimization of \( -\ell(z) \) rather than maximization of \( \ell(z) \).

3. Linear equality constraints in GLLVM

Usually the optimization of a log-likelihood under linear equality constraints is achieved through Lagrange multipliers. However, this procedure does not take advantage of the reduction of dimensionality that is achieved in the parameter vector under certain types of constraints.

We discuss two approaches for handling equality constraints in GLLVM; one is based on the known Lagrange multipliers method (see Section 3.1) and the other one, that is proposed here, reduces the dimension of the optimization procedure under the constrained model (see Section 3.2).
3.1. Constrained optimization via Lagrange multipliers

In order to use the Lagrange multiplier rule in the minimization of \(-\ell(\mathbf{z})\) subject to the constraints given in (1) we need to expand \(\ell(\mathbf{z})\) in a second-order Taylor series about the current approximation \(\mathbf{z}_k\) to the constrained minimum given by

\[
\ell(\mathbf{z}) \approx \ell(\mathbf{z}_k) + \ell'(\mathbf{z}_k)^T(\mathbf{z} - \mathbf{z}_k) + \frac{1}{2}(\mathbf{z} - \mathbf{z}_k)^T\ell''(\mathbf{z}_k)(\mathbf{z} - \mathbf{z}_k).
\]

Then the quadratic approximation to \(\ell(\mathbf{z})\) is minimized subject to the constraints by minimizing the Lagrangian given by

\[
L(\mathbf{z}, \lambda) = \ell'(\mathbf{z}_k)^T(\mathbf{z} - \mathbf{z}_k) + \frac{1}{2}(\mathbf{z} - \mathbf{z}_k)^T\ell''(\mathbf{z}_k)(\mathbf{z} - \mathbf{z}_k) + \lambda^T(V\mathbf{z} - \mathbf{b})
\]

and thus we obtain the next iterate \(\mathbf{z}_{k+1}\) according to

\[
\begin{pmatrix}
\mathbf{z}_{k+1} \\
\lambda
\end{pmatrix} = \begin{pmatrix}
\ell''(\mathbf{z}_k) & V^T \\
V & 0
\end{pmatrix}^{-1}
\begin{pmatrix}
-\ell'(\mathbf{z}_k) \\
\mathbf{b}
\end{pmatrix},
\]

(5)

where \(\ell'(\mathbf{z}_k)\) and \(\ell''(\mathbf{z}_k)\) are the score vector and Hessian matrix for the GLLVM, respectively. The score vector \(\ell'(\mathbf{z}_k)\) of the unconstrained GLLVM is given in (9).

The inversion required in (5) can be accomplished using the following result:

\[
\begin{pmatrix}
\ell''(\mathbf{z}_k) & V^T \\
V & 0
\end{pmatrix}^{-1} = \begin{pmatrix}
\ell''(\mathbf{z}_k)^{-1} & \sqrt{\Omega} V \ell''(\mathbf{z}_k)^{-1} \\
\sqrt{\Omega} V^T & \Omega^{-1} - (\sqrt{V \ell''(\mathbf{z}_k)^{-1} V^T})^{-1}
\end{pmatrix}
\]

which saves computation time and storage (if and only if \(V\) has linearly independent rows \(v_i^T, \ldots, v_p^T\)) where \(\Omega = \ell''(\mathbf{z}_k)^{-1} V^T(\sqrt{V \ell''(\mathbf{z}_k)^{-1} V^T})^{-1}\).

In the above recursion formula, the computation of the Hessian \(\ell''(\mathbf{z}_k)\) is required in each iteration. This may prove to be computationally intensive especially in cases where a large number of model parameters is estimated. An alternative is to substitute an easily computed positive definite approximation \(A_k\) to \(\ell''(\mathbf{z}_k)\). This leads to a quasi-Newton method that attempts to mimic Newton’s method without directly computing the Hessian matrix. The idea is to initiate the algorithm with a positive definite approximation to the Hessian and then perform low rank updates to this matrix. More details about quasi-Newton methods can be found in \(\text{Lange (2004)}\).

3.2. Constrained optimization via a dimension reduction method

Although, the Lagrange multipliers is a general method mainly used in the optimization of the log-likelihood under linear equality constraints, such a procedure does not take advantage of the reduction of dimensionality that can be achieved when certain types of equality constraints are considered. Here, we propose a method for estimating the GLLVM subject to linear equality constraints that reduces the dimension of the estimation problem.

When constraints are imposed, the derivatives of the log-likelihood given in (4) (denoted here as \(\ell_C\)) with respect to the model parameters \(\mathbf{z}_i, i = 1, \ldots, p, l = 0, 1, \ldots, q (l = 0\) for the intercept and \(l \neq 0\) for the factor loadings), using the Gauss–Hermite quadrature rule for approximating the integrals, are

\[
\frac{\partial \ell_C(\mathbf{z})}{\partial \mathbf{z}_i} = -\sum_{r_1, \ldots, r_q} \sum_{s \in i} \left[ r_{st} - N_s b'_s(\theta_s) \right] \left[ \frac{1}{\text{Var}(x_s)} \cdot \sum_{l=0}^q z_{sl} \right],
\]

where

\[
r_{st} = \sum_{m=1}^n x_{sm} h(z_t | x_m),
\]

(6)

(7)
If we impose constraints on the parameters of item \( i \) then the partial derivatives take the form
\[
\frac{\partial \ell}{\partial z_{il}} = - \sum_{t_1, \ldots, t_q} \sum_s \left[ r_{st} - N t b_i' (\theta_s) \right] \frac{1}{u' (\mu_s)} \cdot \left( \sum_{l=0}^q z_{tl} \right),
\]

If we impose constraints on the parameters of different items, then the partial derivatives take the form
\[
\frac{\partial \ell}{\partial z_{il}} = - \sum_{t_1, \ldots, t_q} \left[ r_{st} - N t b_i' (\theta_s) \right] \frac{1}{u' (\mu_s)} \cdot \left( \sum_{l=0}^q z_{tl} \right).
\]

Obviously, in the case of fixed value constraints, (6) equals zero. When linear relationships between the parameters (i.e., negative, additive or proportionality constraints) are imposed, the partial derivatives are easily derived by inserting in front of the components of the summation \( \sum_{s \in \{1, \ldots, p\}} \) or \( \sum_{l=0}^q \) the corresponding coefficients of the parameters in these relationships.

The partial derivatives of \( z_{il} \) for the unconstrained model are
\[
\ell' = \frac{\partial \ell (\mathbf{x})}{\partial z_{il}} = - \sum_{t_1, \ldots, t_q} \left[ r_{st} - N t b_i' (\theta_s) \right] \frac{1}{u' (\mu_s)} z_{tl}, \quad l = 0, 1, \ldots, q, \quad i = 1, \ldots, p,
\]
where \( z_{i0} = 1 \).

According to (6) and (9) the score vector of the constrained model can be expressed in terms of the score vector of the unconstrained model \( \ell' \). This can be done by using an indicator matrix \( \mathbf{A} \) of dimension \( v \times v \) where \( v \) is equal to the total number of model parameters. More specifically, the elements of matrix \( \mathbf{A} \) indicate which parameters are selected to be equal or take fixed values. For example, a value equal to 1 implies a positive equality constraint and \(-1\) a negative one. An example on the construction of the indicator matrix \( \mathbf{A} \) is given in the Appendix. Thus, the score vector \( \ell'_C \) of the constrained model is
\[
\ell'_C = \mathbf{A} \cdot \ell'.
\]

Note that matrix \( \mathbf{A} \) given in (10) is not of full rank due to the equality parameter constraints. The full rank matrix \( \mathbf{A}^C \) obtained by deleting the linearly dependent rows of \( \mathbf{A} \) is of reduced dimension and its multiplication with the vector \( \ell' \) leads to a reduction in the dimension of vector \( \ell'_C \) as well. The dimension reduction is achieved for the type of constraints (1)–(5) given in Section 1. Note that for the case of additive constraints, linear relationships between only two parameters can be assumed. For example, under the additive constraint \( z_{il} + c_1 z_{i'l'} + c_2 z_{i''l''} = 5c_3 \) with \( c_1, c_2, c_3 \in \mathbb{R}, i, i', i'' = 1, \ldots, p \) and \( l, l', l'' = 1, \ldots, q \), the matrix \( \mathbf{A} \) is of full rank and thus no dimension reduction is achieved.

The estimation of the constrained GLLVM can be accomplished with standard numerical methods. Two popular methods for obtaining maximum likelihood estimates are the expectation–maximization (E–M) algorithm (Dempster et al., 1977) and the Newton–Raphson (N–R) algorithm. However, it is common practice in the area of GLLVM to use the E–M since the N–R requires a computationally intensive calculation of the Hessian. The steps of the E–M
We expand that is, a N–R step with the Hessian substituted by a positive definite matrix. A good candidate for the positive definite data log-likelihood computed in the E-step is given in (10). In the M-step we need to solve the non-linear equations \( \frac{\partial \ell_c}{\partial \theta_i} = 0 \). Since these equations do not have a closed form solution a quasi-Newton step is adopted (Lange, 1995). That is, a N–R step with the Hessian substituted by a positive definite matrix. A good candidate for the positive definite matrix is the expected second-order derivative of the expected complete data log-likelihood \( (I_C) \) used in the E-step of the E–M algorithm. This matrix for the GLLVM under positive equality and fixed value constraints is

\[
I_C = \left[ A^C \right]^T \cdot B \cdot A^C, 
\]

where

- \( A^C \) is the indicator matrix of the constraints of dimension smaller than \( A \),
- \( [A^C]^T \) is the transpose of the matrix \( A^C \),
- \( B \) is a block diagonal matrix where each block is the matrix of the expected second-order derivatives of the expected complete data log-likelihood for each item.

The advantages of using the derived matrix \( I_C \) instead of any other positive definite matrix is that the former is properly scaled. Otherwise, the algorithm may have difficulties in the search for a better point at each iteration which inevitably impels the algorithm to many iterations to compensate for the poor initial scaling. However, although the matrix \( I_C \) is of reduced dimension, its derivation requires the computation of \( B \) in each iteration according to (11) which has the same dimension of the unconstrained model. Alternatively, we could update \( I_C \) (or any other positive definite matrix) using quasi-Newton methods (see Section 3.1).

We propose for the estimation of the constrained GLLVM a hybrid algorithm (Baker, 1994) that begins with a moderate number of E–M iterations, as described above, followed by N–R iterations where the \( I_C \) matrix of the last E–M iteration is updated using quasi-Newton methods. The obvious advantages of the hybrid algorithm lie in the quick convergence to the optimum and in dimension reduction of the optimization. We note that although both the Lagrange method and the hybrid algorithm require the computation of the gradient \( \ell' \) of the unconstrained model, the optimization method proposed here will always be of smaller dimension.

### 4. Optimization under inequality constraints

The minimization of a twice continuously differentiable function \( f(\mathbf{x}) \) subject to linear inequality parameter constraints is usually done with the Lagrange multiplier method and the use of slack variables. However, an easier method for this constrained minimization is the adaptive barrier method (Lange, 1994) that we propose to use here for handling the inequality constraints. Let us assume that we want to minimize the minus log-likelihood function \( \ell(\mathbf{x}) \) of a GLLVM subject to the linear inequality constraints \( w_i(\mathbf{x}) = u_i^T \mathbf{x} - d_i \geq 0 \) for \( 1 \leq i \leq l \) where \( u_i \) is the \( i \)th indicator vector of the constraints. According to the adaptive barrier method, instead of minimizing \( \ell(\mathbf{x}) \) we can minimize the surrogate function

\[
R(\mathbf{x} | \mathbf{x}_k) = \ell(\mathbf{x}) - \mu \sum_{i=1}^{l} w_i(\mathbf{x}_k) \ln w_i(\mathbf{x}) - u_i^T \mathbf{x},
\]

for some constant \( \mu > 0 \) where \( \mathbf{x}_k \) belongs to the interior of the region \( U = \{ \mathbf{x} : w_i(\mathbf{x}) > 0 \text{ for all } i \} \).

However, an explicit minimization of \( R(\mathbf{x} | \mathbf{x}_k) \) is not always possible and thus a quadratic expansion is used instead. We expand \( R(\mathbf{x} | \mathbf{x}_k) \) in a second-order Taylor series around the current approximation \( \mathbf{x}_k \) to the constrained minimum

\[
R(\mathbf{x} | \mathbf{x}_k) \approx R(\mathbf{x} | \mathbf{x}_k) + R'(\mathbf{x}_k | \mathbf{x}_k)^T (\mathbf{x} - \mathbf{x}_k) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_k)^T R''(\mathbf{x}_k | \mathbf{x}_k) (\mathbf{x} - \mathbf{x}_k) .
\]

The next iterate is obtained according to

\[
\mathbf{x}_{k+1} = \mathbf{x}_k - R''(\mathbf{x}_k | \mathbf{x}_k)^{-1} R'(\mathbf{x}_k | \mathbf{x}_k) .
\]

Since the computation of the Hessian \( R''(\mathbf{x}_k | \mathbf{x}_k) \) in each iteration is computationally intensive a positive definite approximation to the Hessian \( R''(\mathbf{x}_k | \mathbf{x}_k) \) can be used instead that leads to quasi-Newton methods.
Apart from the minimization of the minus log-likelihood function of a GLLVM under inequality constraints, the minimization with additional equality constraints can be implemented by applying the methods proposed in Section 3.2 to the quadratic approximation of \( R(x|\alpha_k) \) given in (12).

5. Computation of standard errors

Parameter constraints affect the computation of standard errors of the estimated parameters. According to Lange (1999), to calculate the asymptotic covariance matrix of the MLE \( \hat{\alpha} \) with the type of constraints given in (1), we need to re-parameterize the parameter vector \( \alpha \) as follows:

\[
\alpha = \gamma + W\beta,
\]

where \( \gamma \) is a solution of (1), \( W \) is a \( v \times (v - b) \) matrix with \( v - b \) linearly independent columns \( w_1, \ldots, w_{v-b} \) orthogonal to the rows \( v_1, \ldots, v_b \) of \( V \) and \( \beta \) is the parameter vector defined as

\[
\{ \alpha \in \mathbb{R}^v : V\alpha = b \} = \{ \gamma + W\beta : \beta \in \mathbb{R}^{v-b} \}.
\]

The matrix \( W \) can be derived via the formula

\[
W = I_v - V^T (VV^T)^{-1} V,
\]

where \( I_v \) is the \( v \times v \) identity matrix, or alternatively using the Gram–Schmidt method (QR decomposition). If we let \( I(\hat{\alpha}) \) represent either the observed information \( (\ell''(\hat{\alpha})) \) or the expected information \( (E[-\ell''(\hat{\alpha})]) \) of the original parameters, then it is shown that an asymptotic covariance matrix of the estimated parameter vector \( \hat{\alpha} \) with the constraints (1) is given by

\[
Var(\hat{\alpha}) = W(W^T I(\hat{\alpha}) W)^{-1} W^T.
\]

In the presence of linear inequality constraints we calculate the covariance matrix of \( \hat{\alpha} \) by ignoring the inactive constraints and including the active constraints to the existing linear equality constraints. Thus, the constraint matrix \( V \) and its orthogonal \( W \) are of larger dimensions.

6. Hypothesis testing

Statistical comparisons between the unconstrained and constrained model are done with the likelihood ratio test (LRT) applied to nested models. When the null hypothesis includes only equality constraints, it is well known that the LRT statistic is approximated by the \( \chi^2 \) distribution with degrees of freedom equal to the difference between the number of parameters of the models under comparison (see Appendix). When inequality constraints are involved in the null hypothesis, it is proven that the LRT statistic, when \( H_0 \) holds, follows approximately a chi-bar squared distribution (Robertson and Dykstra, 1988). That is a weighted sum of chi-squared distributions where the weights can be estimated via Monte Carlo methods (Dardanoni and Forcina, 1998). Alternatively, the \( p \)-value of the statistic can be computed via parametric bootstrapping (Ritov and Gilula, 1993). Also, Cysneirosa and Paulab (2005) discuss the problem of testing equality and inequality constraints in regression analysis-type models.

7. Simulation study

In this section, we evaluate through a simulation study the performance and efficiency of the proposed methodology in estimating a GLLVM subject to equality constraints versus the commonly used Lagrange multiplier method. As we have discussed in Section 3.2, the advantage of our approach over the Lagrange multiplier method is the considerable dimension reduction that is achieved in the score vector and the Hessian matrix. Reducing the dimensionality of the optimization can have a direct impact on the efficiency of the algorithm and this can be assessed by counting the number of floating points operations (FLOPs) that are required in each iteration. The main effect of the dimensionality reduction is in the inversion of the Hessian that has to be done in each iteration. In fact, it is known that the matrix
Table 1
Parameter estimates, standard errors in brackets and bias from the 10,000 simulated data using the dimension reduction method

<table>
<thead>
<tr>
<th>Item</th>
<th>True values</th>
<th>Dimension reduction method—Parameter estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \theta_0 )</td>
<td>( \theta_1 )</td>
</tr>
<tr>
<td>1</td>
<td>0.17</td>
<td>−0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>1.49</td>
</tr>
<tr>
<td>3</td>
<td>0.14</td>
<td>1.07</td>
</tr>
<tr>
<td>4</td>
<td>0.08</td>
<td>2.25</td>
</tr>
<tr>
<td>5</td>
<td>2.35</td>
<td>1.53</td>
</tr>
<tr>
<td>6</td>
<td>−0.23</td>
<td>1.25</td>
</tr>
<tr>
<td>7</td>
<td>0.82</td>
<td>−0.39</td>
</tr>
<tr>
<td>8</td>
<td>−0.02</td>
<td>2.47</td>
</tr>
<tr>
<td>9</td>
<td>0.43</td>
<td>0.63</td>
</tr>
<tr>
<td>10</td>
<td>0.43</td>
<td>1.40</td>
</tr>
</tbody>
</table>

Table 2
Parameter estimates, standard errors in brackets and bias from the 10,000 simulated data sets using the Lagrange multipliers method

<table>
<thead>
<tr>
<th>Item</th>
<th>True values</th>
<th>Lagrange multipliers method—Parameter estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \theta_0 )</td>
<td>( \theta_1 )</td>
</tr>
<tr>
<td>1</td>
<td>0.17</td>
<td>−0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>1.49</td>
</tr>
<tr>
<td>3</td>
<td>0.14</td>
<td>1.07</td>
</tr>
<tr>
<td>4</td>
<td>0.08</td>
<td>2.25</td>
</tr>
<tr>
<td>5</td>
<td>2.35</td>
<td>1.53</td>
</tr>
<tr>
<td>6</td>
<td>−0.23</td>
<td>1.25</td>
</tr>
<tr>
<td>7</td>
<td>0.82</td>
<td>−0.39</td>
</tr>
<tr>
<td>8</td>
<td>−0.02</td>
<td>2.47</td>
</tr>
<tr>
<td>9</td>
<td>0.43</td>
<td>0.63</td>
</tr>
<tr>
<td>10</td>
<td>0.43</td>
<td>1.40</td>
</tr>
</tbody>
</table>

inversion is equivalent to \( O(3^3) \) FLOPs, where \( v \) denotes the matrix dimension. Thus, comparing the FLOPs required for this computation by the two methods we can get an idea on their relative efficiency. However, a fair comparison between numerical approaches should also take into account the required number of iterations until convergence.

7.1. Setup of the simulation study—results

In the simulation study we consider the following setting: a sample of \( n = 1000 \) students are assumed to have taken a test designed to measure two aspects of their mathematical ability and a correct or wrong answer to these questions has been recorded. The test consists of 10 binary questions, with the first and the second, the ninth and the tenth items having the same intercepts (easiness level), while the first five items are assumed to have the same factor loadings on both factors (discrimination power). We are thus interested in fitting a two-factor logit model to the 10 binary items, subject to seven sets of equality constraints. We have simulated 10,000 data sets, under the above setting and the constrained model is fitted using both the proposed methodology and the Lagrange multipliers method. Their efficiency in estimating the model parameters is evaluated. The dimension of the Hessian matrix in our approach is 23 \( \times \) 23, whereas in the Lagrange multipliers method is 30 \( \times \) 30. This implies that reducing the dimension of the Hessian by seven will result in reducing the FLOPs required for its inversion from \( O(30^3) \) to \( O(23^3) \).

The results of the simulation study are presented in Tables 1 and 2.

We observe that both methods produce consistent estimates for the constrained two-factor model. In addition, in order to get a rough estimate of the reduction in FLOPs achieved with our method, we have recorded for each data set the number of times the Hessian is inverted until convergence, under both approaches, and multiplied it with the
required FLOPs per iteration. Thus, using our approach the FLOPs are, on average, reduced from $40.5 \times O(30^3)$ to $3.5 \times O(23^3)$. In addition, using the R function `system.time()` we report here in system CPU the time required by the two methods to perform the optimization for the above simulations. The times required are 7.23 h and 11.68 min system CPU and 25.48 min system CPU for the proposed and the Lagrange method, respectively. The software used was R, version 2.2.1 for Windows XP on a Pentium 4, 3.2 GHz with 2GB ram.

8. Application

To illustrate the methodology developed for both equality and inequality constraints data from the 1996 British Social Attitudes Survey (BSA)\(^1\) and data from the Law School Admission Test (LSAT) given in Bock and Lieberman (1970) will be used. For the second example the library \texttt{ltm} (Rizopoulos and Moustaki, 2006) is used to estimate the unconstrained models while all the other models were fitted with functions written by the authors in S/R.

8.1. Example I

The first data set consists of five binary items, designed to measure satisfaction with the National Health Service in respondents’ area and more specifically with services provided by general practitioners (GP). The items asked to 841 individuals concerned whether the National Health Service in your area is, on the whole, satisfactory or in need of improvement. The five items have been analyzed in Moustaki (2003) with an unconstrained latent variable model. We wanted to test whether intercepts and factor loadings are the same for men and women. We started the analysis by fitting the unconstrained model. The estimated model parameters with their estimated standard errors are given in Table 3. Standard errors are calculated via forward difference approximation to the Hessian matrix at the estimated parameters. Factor loadings are found to be of similar magnitude for both men and women. This implies that an assumption of item invariance between the two groups is plausible.

We continued the analysis by allowing the intercepts and factor loadings of the two groups to be equal. An one-factor model under equality constraints on all the parameters between the two groups was fitted. The estimated model parameters with estimated standard errors are given in Table 4. The standard errors of the constrained parameters are calculated using (13) and they have all been reduced.

A comparison between the constrained and the unconstrained model is shown in Table 5. Both the LRT, the BIC and AIC criteria suggest that the constrained model is preferred.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Item</th>
<th>$\hat{z}_{i0}$</th>
<th>s.e. ($\hat{z}_{i0}$)</th>
<th>$\hat{z}_{i1}$</th>
<th>s.e. ($\hat{z}_{i1}$)</th>
<th>st ($\hat{z}_{i1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>1</td>
<td>0.68</td>
<td>0.16</td>
<td>1.48</td>
<td>0.26</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.89</td>
<td>0.34</td>
<td>2.50</td>
<td>0.50</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.47</td>
<td>0.23</td>
<td>1.83</td>
<td>0.33</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.89</td>
<td>0.30</td>
<td>2.19</td>
<td>0.41</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.79</td>
<td>0.23</td>
<td>1.51</td>
<td>0.28</td>
<td>0.83</td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
<td>0.49</td>
<td>0.16</td>
<td>2.00</td>
<td>0.29</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.62</td>
<td>0.27</td>
<td>2.80</td>
<td>0.47</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.72</td>
<td>0.24</td>
<td>2.26</td>
<td>0.34</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.56</td>
<td>0.36</td>
<td>2.70</td>
<td>0.45</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.90</td>
<td>0.18</td>
<td>1.21</td>
<td>0.20</td>
<td>0.77</td>
</tr>
</tbody>
</table>

---

Table 4
Parameter estimates, standard errors, standardized factor loadings for the constrained one-factor model, BSA data

<table>
<thead>
<tr>
<th>Gender</th>
<th>Item</th>
<th>$\hat{\theta}_0$</th>
<th>s.e. ($\hat{\theta}_0$)</th>
<th>$\hat{\theta}_1$</th>
<th>s.e. ($\hat{\theta}_1$)</th>
<th>st ($\hat{\theta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female/male</td>
<td>1</td>
<td>0.57</td>
<td>0.11</td>
<td>1.74</td>
<td>0.19</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.73</td>
<td>0.21</td>
<td>2.65</td>
<td>0.34</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.61</td>
<td>0.17</td>
<td>2.08</td>
<td>0.24</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.22</td>
<td>0.23</td>
<td>2.41</td>
<td>0.30</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.84</td>
<td>0.14</td>
<td>1.32</td>
<td>0.16</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 5
Comparison between the constrained and the unconstrained model, BSA data

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-likelihood</th>
<th>LRT value</th>
<th>d.f.</th>
<th>p-value</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constrained</td>
<td>$-2119.473$</td>
<td>4258.95</td>
<td>10</td>
<td>0.091</td>
<td>4286.62</td>
<td>4653.64</td>
</tr>
<tr>
<td>Unconstrained</td>
<td>$-2111.311$</td>
<td>16.324</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6
Parameter estimates, standard errors, standardized factor loadings for the unconstrained one-factor model, LSAT data

<table>
<thead>
<tr>
<th>Item</th>
<th>$\hat{\theta}_0$</th>
<th>s.e. ($\hat{\theta}_0$)</th>
<th>$\hat{\theta}_1$</th>
<th>s.e. ($\hat{\theta}_1$)</th>
<th>st ($\hat{\theta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.77</td>
<td>0.21</td>
<td>0.83</td>
<td>0.26</td>
<td>0.64</td>
</tr>
<tr>
<td>2</td>
<td>0.99</td>
<td>0.09</td>
<td>0.72</td>
<td>0.19</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.08</td>
<td>0.89</td>
<td>0.23</td>
<td>0.66</td>
</tr>
<tr>
<td>4</td>
<td>1.29</td>
<td>0.10</td>
<td>0.69</td>
<td>0.19</td>
<td>0.57</td>
</tr>
<tr>
<td>5</td>
<td>2.05</td>
<td>0.14</td>
<td>0.66</td>
<td>0.21</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Table 7
Parameter estimates, standard errors, standardized factor loadings for the constrained one-factor model, LSAT data

<table>
<thead>
<tr>
<th>Item</th>
<th>$\hat{\theta}_0$</th>
<th>s.e. ($\hat{\theta}_0$)</th>
<th>$\hat{\theta}_1$</th>
<th>s.e. ($\hat{\theta}_1$)</th>
<th>st ($\hat{\theta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.76</td>
<td>0.20</td>
<td>0.80</td>
<td>0.25</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>1.17</td>
<td>0.07</td>
<td>1.02</td>
<td>0.24</td>
<td>0.71</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.07</td>
<td>0.80</td>
<td>0.21</td>
<td>0.62</td>
</tr>
<tr>
<td>4</td>
<td>1.17</td>
<td>0.07</td>
<td>0.54</td>
<td>0.14</td>
<td>0.47</td>
</tr>
<tr>
<td>5</td>
<td>2.05</td>
<td>0.13</td>
<td>0.65</td>
<td>0.20</td>
<td>0.54</td>
</tr>
</tbody>
</table>

8.2. Example II

The second data set consists of five questions, addressed to 1000 test-takers, designed to measure a single latent ability. This example is chosen for illustrating inequality constraints. Since no information is known with respect to the intercepts or factor loadings inequality constraints will be imposed only for illustrative purposes. We test whether the second item has smaller intercept than the fourth (that implies that the second item is more difficult than the fourth).

We started the analysis by fitting the unconstrained model. The estimated model parameters with estimated standard errors are given in **Table 6**. The factor loadings are of similar magnitude.

We continued by constraining the intercept parameter of the second item to be smaller than that of the fourth item. The estimated model parameters with estimated standard errors are given in **Table 7**. We see that the standard errors of the estimated parameters for some of the items have been slightly increased.

A comparison between the constrained and the unconstrained model is based on the LRT. The $p$-value of this test is derived via parametric bootstrapping. In particular, we simulated 100 samples from the model estimated under the null hypothesis. Each of these samples was fitted under both hypotheses and the value of the LRT was recorded.
Table 8
Comparison between the constrained and the unconstrained model, LSAT data

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-Likelihood</th>
<th>Observed LRT value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constrained</td>
<td>−2468.948</td>
<td>4.562</td>
<td>0.614</td>
</tr>
<tr>
<td>Unconstrained</td>
<td>−2466.667</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The number of times the value of the LRT exceeded the value of the LRT in the original sample constitutes the p-value of the test. The result of the parametric bootstrap is given in Table 8.

The high p-value of the LRT indicates that the most preferable model is the constrained one.

9. Conclusion

In this paper, we discussed maximum likelihood estimation for GLLVM subject to general linear equality and inequality constraints. In addition to known methods for handling these constraints, we presented an efficient estimation method for fitting GLLVM subject to general linear equality constraints that achieves a considerable reduction of dimensionality especially when many items and many equality constraints are considered. The merits of the proposed method can be seen especially in the multigroup analysis where the measurement invariance property of the items across groups needs to be tested as in the first example of Section 8. Furthermore, issues concerning the computation of standard errors for the estimated parameters of the constrained models and the estimation of the p-value for the LRT were discussed.

Appendix

The first derivative of the constrained log-likelihood defined in (6) is given here.

The unknown model parameters \( \mathbf{z} \) are found in the canonical parameter \( \theta \) and the specific functions \( b_i (\theta_i) \) and \( c_i (\phi_j, x_m) \) given in (4). When we differentiate the log-likelihood given in (4) with respect to the model parameters \( z_{il}, i = 1, \ldots, p, l = 0, 1, \ldots, q \) (i.e. the constant \( l = 0 \) and factor loadings \( l \neq 0 \)) we get

\[
\frac{\partial \ell (\mathbf{z})}{\partial z_{il}} = \sum_{m=1}^{n} \frac{1}{f (\mathbf{x}_m)} \sum_{t_1, \ldots, t_q} h (\mathbf{z}_t) \frac{\partial}{\partial z_{il}} \prod_{i=1}^{p} g (x_{mi} | \mathbf{z}_t) \\
= \sum_{m=1}^{n} \sum_{t_1, \ldots, t_q} h (\mathbf{z}_t | \mathbf{x}_m) \frac{\partial}{\partial z_{il}} \sum_{i \in I} \left\{ \frac{x_{sm} \theta_s - b_s (\theta_s)}{\phi_s} \right\}. \tag{14}
\]

We see from (14) that the differentiation of \( \ell (\mathbf{z}) \) is done separately for each item \( i \). Therefore when an equality constraint is imposed to the parameters of items \( i \) and \( j \) the derivation of \( \ell (\mathbf{z}) \) will result from summing up the derivatives of these items.

The calculation of the term \( \{\partial / \partial z_{sl}\} \{ (x_{sm} \theta_s - b_s (\theta_s)) / \phi_s \} \) results in

\[
\frac{\partial}{\partial z_{sl}} \left\{ \frac{x_{sm} \theta_s - b_s (\theta_s)}{\phi_s} \right\} = \left\{ \frac{\partial x_{sm} \theta_s - b_s (\theta_s)}{\phi_s} \right\} \frac{\partial \theta_s}{\partial \theta_s} \frac{\partial c_s}{\partial \mu_s} \frac{\partial \eta_s}{\partial \zeta_{sl}} \\
= \left\{ \frac{x_{sm} - b'_s (\theta_s)}{\phi_s} \right\} \frac{\partial \theta_s}{\partial \theta_s} \frac{\partial c_s}{\partial \mu_s} \frac{\partial \eta_s}{\partial \zeta_{sl}}. \tag{15}
\]

Then, by substituting (15) in (14) we get

\[
\frac{\partial \ell}{\partial z_{sl}} = \sum_{m=1}^{n} \sum_{t_1, \ldots, t_q} h (\mathbf{z}_t | \mathbf{x}_m) \sum_{s} \left\{ \frac{x_{sm} - b'_s (\theta_s)}{\phi_s} \right\} \frac{\partial \theta_s}{\partial \theta_s} \frac{\partial c_s}{\partial \mu_s} \frac{\partial \eta_s}{\partial \zeta_{sl}}.
\]
The dimension of matrix and 

Let us assume that the dependencies between three binary items 

In addition, let us assume that we are interested in testing simultaneously the following hypotheses: 

If we consider model (17) without constraints the vector of partial derivatives 

where 

\[
\frac{\partial \theta_s}{\partial \mu_s} = \left[ \frac{\partial \mu_s}{\partial \theta_s} \right]^{-1} = \frac{1}{b' (\theta_s)} , \quad \mu_s = b' (\theta_s) , \\
\frac{\partial \mu_s}{\partial \eta_s} = \left[ \frac{\partial \eta_s}{\partial \mu_s} \right]^{-1} = \frac{1}{u' (\mu_s)} , \quad \eta_s = u (\mu_s) , \\
\frac{\partial \eta_s}{\partial \alpha_{sl}} = \frac{\partial}{\partial \alpha_{sl}} \left( x_{s0} + \sum_{j=1}^{q} x_{sl} z_{jt} \right) , \quad (16)
\]

and \( r_{st}, N_i \) are given by (7) and (8).

Construction of the constraints matrix \( A \). The construction of the matrix \( A \) is presented using an artificial example. Let us assume that the dependencies between three binary items \( x_1, x_2, x_3 \) can be accounted for by two independent latent variables \( z_1 \) and \( z_2 \). Thus, model (3) is of the form:

\[
\begin{align*}
\logit \pi_1 &= x_{10} + x_{11} z_1 + x_{12} z_2, \\
\logit \pi_2 &= x_{20} + x_{21} z_1 + x_{22} z_2, \\
\logit \pi_3 &= x_{30} + x_{31} z_1 + x_{32} z_2. \\
\end{align*}
\]

(17)

In addition, let us assume that we are interested in testing simultaneously the following hypotheses:

\[
H_0 : \begin{cases} 
\pi_{11} = \pi_{12} = \pi_{21}, \\
\pi_{22} = \pi_{32}, 
\end{cases} \quad H_1 : \alpha_{ij} \text{ are not constrained. (18)}
\]

If we consider model (17) without constraints the vector of partial derivatives \( \ell' \) is given by (9) for \( i = 1, 2, 3 \) and \( j = 1, 2 \). However, under the equality parameter constraints (18) the vector of the partial derivatives \( \ell'_C \) takes the form:

\[
\ell'_C = \begin{bmatrix}
\sum_{t_1, t_2} D_1 \\
\sum_{t_1, t_2} D_1 (z_{1t} + z_{2t}) + \sum_{t_1, t_2} D_2 z_{1t} \\
\sum_{t_1, t_2} D_1 (z_{1t} + z_{2t}) + \sum_{t_1, t_2} D_2 z_{1t} \\
\sum_{t_1, t_2} D_2 z_{1t} + \sum_{t_1, t_2} D_2 z_{2t} \\
\sum_{t_1, t_2} D_2 z_{1t} + \sum_{t_1, t_2} D_2 z_{2t} \\
\sum_{t_1, t_2} D_3 z_{1t} + \sum_{t_1, t_2} D_3 z_{2t} \\
\sum_{t_1, t_2} D_3 z_{1t} + \sum_{t_1, t_2} D_3 z_{2t}
\end{bmatrix}, \quad (19)
\]

where \( D_i = [(r_{it} - N_i \cdot b' (\theta_i))/\text{Var} (x_i)] 1/\mu' (\mu_i) \) and \( i = 1, 2, 3 \). We observe that vector (19) can be written as the product of vector \( \ell' \) and matrix \( A \) according to (10). The matrix \( A \) of the above constrained model is

\[
A = \begin{bmatrix}
x_{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
x_{11} & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
x_{12} & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
x_{20} & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
x_{21} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
x_{22} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
x_{30} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
x_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
x_{32} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 
\end{bmatrix}
\]

The dimension of matrix \( A \) equals the number of model parameters. We have named the rows and columns after the parameters to be estimated. Values equal to 1 indicate that the two corresponding parameters are equal. In our example,
parameter \( z_{10} \) does not equal to any of the other parameters of the model, therefore, we set only the first element of the first row to 1 and 0 everywhere else. The second row corresponds to parameter \( z_{11} \). According to the constraints, \( z_{11} = z_{12} = z_{21} \), we put into the second, the third and the fifth column of the second row the value 1. The same procedure is followed for the rest of the parameters. The elements of the main diagonal will always be 1 except from the parameters that are constrained to be equal at a fixed value and then it will take the value 0.

**Degrees of freedom.** For the reasons presented in Section 1 we might be interested in checking hypotheses of the form:

\[
\begin{align*}
\text{H}_0 : & \ z_{31} = z_{42} = 0, \\
\text{H}_1 : & \ z_{ij} \text{ are not constrained.}
\end{align*}
\]

In other words, we may be interested in checking whether the loading of the first factor for the fifth item equals the loading of the second factor for the fifth item or whether the loading of the first factor for the third item and the loading of the second factor for the fourth item equal zero or whether the factor loadings for all categories of an item are equal. That implies that the number of the unknown model parameters \( v \) is reduced under the various constraints. For example, when the observed variables \( x \) are binary, the number of parameters to be estimated for the unconstrained model are the \( p \) intercepts and the \( p \times q \) factor loadings, a total of \( v = p \cdot (q + 1) \) unknown parameters. This result holds provided that we are interested in marginal inference. If we impose \( c_1 \) sets of equality constraints, such that each of these sets contains \( d_1 \) parameters, and \( c_2 \) sets of fixed value constraints, such that each of these sets contains \( d_2 \) parameters, then the number of the different parameters in the constrained GLLVM that need to be estimated is \( v - \sum_{i=1}^{c_1} (d_1i - 1) - \sum_{i=1}^{c_2} d_2i \). For the other members of the exponential family the number of degrees of freedom is defined analogously.

**References**


