Factor Analysis of Ordinal Variables with Full Information Maximum Likelihood

Karl G Jöreskog
Uppsala University

Irini Moustaki
Athens University of Economics and Business

20 March 2006

The basic idea of factor analysis is the following. For a given set of manifest variables $x_1, \ldots, x_p$ one wants to find a set of latent variables $\xi_1, \ldots, \xi_k$, fewer in number than the manifest variables, that contain essentially the same information. The latent variables are supposed to account for the dependencies among the manifest variables in the sense that if the latent variables are held fixed, the manifest variables would be independent. Classical factor analysis assumes that both the manifest and the latent variables are continuous variables and is usually carried out by factor analyzing the sample covariance or correlation matrix of the manifest variables. There is a long history of methods for fitting factor models to a correlation or covariance matrix, see *e.g.*, Jöreskog (2006)

In practice, the manifest variables are often ordinal. However, ordinality is most often ignored and numbers such as 1, 2, 3, 4, representing ordered categories, are treated as numbers having metric properties, a procedure which is incorrect in several ways. A model for factor analysis of ordinal variables must specify the probability of each response pattern as a function of $\xi_1, \xi_2, \ldots, \xi_k$:

$$Pr(x_1 = a_1, x_2 = a_2, \ldots, x_p = a_p \mid \xi_1, \xi_2, \ldots, \xi_k) = f(\xi_1, \xi_2, \ldots, \xi_k) \,, \tag{1}$$

where $a_1, a_2, \ldots, a_p$ represent the different response categories of $x_1, x_2, \ldots, x_p$, respectively.

Jöreskog (2002-2005) describes methods for estimating factor analysis models and structural equation models for ordinal variables using only univariate and bivariate information. These methods are based on the assumption of underlying normality. Here we consider full information maximum likelihood (FIML) methods. The full information methods proposed and discussed in the literature differ in terms of the item response function used and the method used to compute the estimates. For example, Muraki (1990) and Moustaki (2003) use the logistic response function and Muraki & Carlson (1995) use the normal ogive response function. Both of these approaches use an EM algorithm described first by Bock & Aitkin (1981) and recently in a more general way in Bartholomew & Knott (1999, pp. 80-83) to obtain the maximum likelihood estimates.

This paper continues the work of Jöreskog & Moustaki (2001) and describes a general approach to factor analysis of ordinal variables and its implementation in LISREL. With this approach one can use either the normal ogive response function, here called NOR, or the logistic response function, here called POM. Both models can be estimated by full information maximum likelihood

(FIML). Our approach differs from other approaches in that we obtain the maximum likelihood estimates of the parameters by direct optimization of the likelihood function.

Section 1 explains what we mean by full information and how the data is organized for efficient computation. Section 2 describes the two models NOR and POM and sections 3 and 4 describe the FIML method of estimation. Section 5 defines standardized and unstandardized parameters. Section 6 discusses measurement of fit. The LISREL implementation is described in section 7. An example is given in section 8.

1 Full Information

Multivariate categorical data can be organized in several ways without loosing any information. Various ways of looking at the data are illustrated here using six ordinal political efficacy variables previously described in Aish & Jöreskog (1990), Jöreskog & Moustaki (2001), and Jöreskog (2002-2005). These data are the responses to the following statements

NOSAY People like me have no say in what the government does

VOTING Voting is the only way that people like me can have any say about how the government runs things

COMPLEX Sometimes politics and government seem so complicated that a person like me cannot really understand what is going on

NOCARE I don't think that public officials care much about what people like me think

TOUCH Generally speaking, those we elect to Congress in Washington lose touch with the people pretty quickly

INTEREST Parties are only interested in people's votes but not in their opinions

The ordered categories are

1 agree strongly (**AS**)

2 agree (**A**)

3 disagree (**D**)

4 disagree strongly (**DS**)

Here we use the 1554 cases in the USA sample with complete responses to all six items.

The most common way of organizing the data is as a data matrix of 1554 rows and 6 columns. Each row of this matrix represent the responses of an individual in the sample.

With six variables and four categories of each variable there are 4096 possible response patterns. So the same data can also be viewed as a $4 \times 4 \times 4 \times 4 \times 4 \times 4$ contingency table. Obviously, many cells in this contingency table are empty. In fact, a closer screening of the data reveals that there are only 476 distinct response patterns occurring in the sample. Thus, the total number of empty cells are $4096 - 476 = 3620$. We refer to the ratio between 476 and 4096 as the *Coverage Ratio (CR)*:

$$CR = \frac{476}{4096} = 0.116 \ . \tag{2}$$

Thus only 11.6% of the response patterns are represented in the sample. With a small *Coverage Ratio* there are many response patterns that are not chosen. In such cases there will not be much information lost if one collapses categories. For example, one may collapse the two agree categories into one category and the two disagree categories into another category. We shall pursue this idea in section 8 where we analyze the data.

CR varies considerably from one data set to another. Data sets with larger values of CR give a better representation of the set of all possible response patterns.

A third way of organizing the data is by listing all distinct response patterns together with its frequency of occurrence. PRELIS automatically produces such a FREQ file for ordinal variables. This FREQ file is used in the estimation procedure. In the example, the FREQ file has 476 lines. The first ten lines representing the ten most common response patterns are as follows

FREQUENCY	NOSAY	VOTING	COMPLEX	NOCARE	TOUCH	INTEREST
97	A	A	A	A	A	A
70	D	D	A	D	D	D
49	D	A	A	A	A	A
45	D	D	D	D	D	D
45	D	D	A	A	A	A
40	D	A	A	D	D	D
32	D	D	A	D	A	A
31	D	D	A	D	A	D
25	A	A	AS	A	A	A
23	AS	AS	AS	AS	AS	AS

This shows that the most common response pattern (97 cases) is to agree to all items. The second most common response pattern (70 cases) is to disagree to all but COMPLEX. There are 23 cases which agree strongly to all six items.

2 Models

Bartholomew & Knott (1999) use the term latent variable models in a general setting. There are two sets of variables, p manifest variables $\mathbf{x}$ and k latent variables $\boldsymbol{\xi}$ with a joint distribution. If the marginal distribution $h(\boldsymbol{\xi})$ of $\boldsymbol{\xi}$ and the conditional distribution $g(\mathbf{x} \mid \boldsymbol{\xi})$ of $\mathbf{x}$ for given $\boldsymbol{\xi}$ exist, then the marginal distribution $f(\mathbf{x})$ of $\mathbf{x}$ must be

$$f(\mathbf{x}) = \int h(\boldsymbol{\xi}) g(\mathbf{x} \mid \boldsymbol{\xi}) d\boldsymbol{\xi} \ . \tag{3}$$

This is a tautology in the sense that it always holds if the distributions exist. However, the idea of factor analysis and latent variable models is that the manifest variables should be independent for given latent variables, *i.e.*, the latent variables should account for all dependencies among the manifest variables. Thus,

$$g(\mathbf{x} \mid \boldsymbol{\xi}) = \prod_{i=1}^{p} g(x_i \mid \boldsymbol{\xi}) \ , \tag{4}$$

so that

$$f(\mathbf{x}) = \int h(\boldsymbol{\xi}) \prod_{i=1}^{p} g(x_i \mid \boldsymbol{\xi}) d\boldsymbol{\xi} \ . \tag{5}$$

To specify a latent variable model, one must therefore specify $h(\boldsymbol{\xi})$ and $g(x_i \mid \boldsymbol{\xi})$ for all i. The latter distribution may be different for different i.

The manifest variables may be continuous or categorical and the latent variables may be continuous or categorical. Thus there may be four classes of latent variable models as shown in Table 1. For a review of various models and approaches, see Moustaki (2006).

Table 1: Classification of Latent Variable Models

	Manifest Variables	
Latent Variables	Continuous	Categorical
Continuous	A: Factor Analysis Models	B: Latent Trait Models
Categorical	C: Latent Profile Models	D: Latent Class Models

Class A is the classical factor analysis model. Assuming normality, one takes $\boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{I})$. The model for normally distributed manifest variables is obtained by taking $\mathbf{x} \mid \boldsymbol{\xi} \sim N(\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\xi}, \boldsymbol{\Psi})$. This implies $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi})$.

Moustaki & Knott (2000) considered Class B with distributions $g(x_i \mid \boldsymbol{\xi})$ chosen from the exponential family. The case considered here is when the manifest variables are binary or ordinal.

Let x_i be an ordinal variable with $m_i \geq 2$ ordered categories $a_i = 1, 2, \ldots, m_i$. The models are defined by the cumulative response function

$$Pr\{x_i \leq s \mid \boldsymbol{\xi}\} = F(\alpha_s^{(i)} - \sum_{j=1}^{k} \beta_{ij}\xi_j) \; , \tag{6}$$

where F is a distribution function and

$$-\infty = \alpha_0^{(i)} < \alpha_1^{(i)} < \alpha_2^{(i)} \cdots < \alpha_{m_i-1}^{(i)} < \alpha_{m_i}^{(i)} = \infty \; .$$

In factor analysis terminology the $\alpha_s^{(i)}$ are intercept terms or threshold parameters and the β_{ij} are factor loadings.

The motivation for the negative sign in (6) is as follows. If β_{ij} is positive, the probability on the left hand side of (6) decreases if ξ_j increases, which is the same as to say that the probability of a respons in a category larger than s on x_i increases with ξ_j. Hence, x_i and ξ_j are measured in the same direction.

In principle, $F(t)$ can be any distribution function defined for $-\infty < t < +\infty$. The only choices of F considered here are the normal (NOR) and the logistic (POM) distribution functions (see Jöreskog & Moustaki, 2001),

$$\mathsf{NOR}: \; F(t) = \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \; , \tag{7}$$

$$\mathsf{POM}: \; F(t) = \Psi(t) = \frac{e^t}{1 + e^t} \; . \tag{8}$$

There are $M = \prod_i^p m_i$ possible response patterns. The model defines the probability of all these response patterns. Let $\mathbf{x}_r = (x_1 = a_1, x_2 = a_2, \ldots, x_p = a_p)$ be a general response pattern, where $a_i = 1, 2, \ldots, m_i$ and let

$$\pi_{a_i}^{(i)}(\boldsymbol{\xi}) = F(\alpha_{a_i}^{(i)} - \sum_{j=1}^{k} \beta_{ij}\xi_j) - F(\alpha_{a_i-1}^{(i)} - \sum_{j=1}^{k} \beta_{ij}\xi_j) \tag{9}$$

be the conditional probability that variable x_i falls in the ordered category a_i. Then the conditional probability of $\mathbf{x}_r$ is

$$\pi_r(\boldsymbol{\xi}) = \prod_{i=1}^{p} \pi_{a_i}^{(i)}(\boldsymbol{\xi}) = \prod_{i=1}^{p} [F(\alpha_{a_i}^{(i)} - \sum_{j=1}^{k} \beta_{ij}\xi_j) - F(\alpha_{a_i-1}^{(i)} - \sum_{j=1}^{k} \beta_{ij}\xi_j)] \ . \tag{10}$$

For exploratory factor analysis we take $\boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{I})$, so that $h(\boldsymbol{\xi}) = \prod_{j=1}^{k} \phi(\xi_j)$. The unconditional probability π_r is

$$\pi_r(\boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \pi_r(\boldsymbol{\xi}) h(\boldsymbol{\xi}) d\boldsymbol{\xi} \ , \tag{11}$$

where $\int$ is a k-dimensional multiple integral evaluated by Gauss-Hermite quadrature. The probability π_r is a function of the parameter vector $\boldsymbol{\theta}$ consisting of all the intercepts $\alpha_s^{(i)}$ and all the factor loadings β_{ij}.

3 Estimation

Let n_r be the frequency of occurrence of the response pattern $\mathbf{x}_r$ and let $p_r = n_r/N$, where N is the sample size. The corresponding probability π_r is defined by (11). The logarithm of the likelihood function is

$$\ln L = \sum_r n_r \ln \pi_r(\boldsymbol{\theta}) = N \sum_r p_r \ln \pi_r(\boldsymbol{\theta}) \ , \tag{12}$$

where the sum runs over all response patterns occurring in the sample, *i.e.*, over all r with $n_r > 0$. Instead of maximizing (12) it is convenient to minimize the fit function, see Jöreskog & Moustaki (2001),

$$F(\boldsymbol{\theta}) = \sum_r p_r [\ln p_r - \ln \pi_r(\boldsymbol{\theta})] = \sum_r p_r \ln[p_r/\pi_r(\boldsymbol{\theta})] \tag{13}$$

This function is non-negative and equals zero only when there is a perfect fit, *i.e.*, when $p_r = \pi_r$ for all r. The minimum value of F is the likelihood ratio chi-square statistic for testing the model against the alternative that the π_r only satisfies $\pi_r > 0$, for all r, and $\sum_r \pi_r = 1$.

4 Minimization Algorithm

To minimize the fit function the gradient vector and the information matrix are needed. The gradient vector is

$$\partial F(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = -\sum_r [p_r/\pi_r(\boldsymbol{\theta})] \partial \pi_r(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \ , \tag{14}$$

and the Hessian matrix is

$$\partial^2 F/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = \sum_r \{-[p_r/\pi_r(\boldsymbol{\theta})]\, \partial^2\pi_r(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' \tag{15}$$

$$+ \quad [p_r/\pi_r^2(\boldsymbol{\theta})]\partial\pi_r(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\ \partial\pi_r(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'\} \ . \tag{16}$$

Note that under the model,

$$E(p_r) = \pi_r(\boldsymbol{\theta}) \ .$$

Since

$$\sum \pi_r(\boldsymbol{\theta}) = 1 \ ,$$

we have

$$\sum \partial\pi_r(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0 \ ,$$

$$\sum \partial\pi_r^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = 0 \ ,$$

Therefore, the information matrix is

$$E(\boldsymbol{\theta}) = -E[\partial^2 \ln L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')] = E[\partial^2 F(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'] = \sum[1/\pi_r(\boldsymbol{\theta})]\partial\pi_r(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\ \partial\pi_r(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' \ . \tag{17}$$

Let $\boldsymbol{\theta}^{(0)}$ be a set of starting values. A Fisher scoring type minimization algorithm generates successive points $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$, …, in the parameter space such that

$$F[\boldsymbol{\theta}^{(s+1)}] < F[\boldsymbol{\theta}^{(s)}] \ .$$

For $s = 0, 1, 2, \ldots$, let $\mathbf{g}^{(s)}$ be the gradient vector and let $\mathbf{E}^{(s)}$ be the inverse of the information matrix at $\boldsymbol{\theta}$=$\boldsymbol{\theta}^{(s)}$. Furthermore, let $\alpha^{(s)}$ be a sequence of positive scalars converging to 0. Then the minimization algorithm is

$$\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)} + \alpha^{(s)}\mathbf{E}^{(s)}\mathbf{g}^{(s)} \ . \tag{18}$$

The sequence $\alpha^{(s)}$ can be chosen so that the function is decreased in each iteration.

5 Standardized and Unstandardized Parameters

In previous sections the model has been formulated in terms of *unstandardized parameters* $\alpha_a^{(i)}$ and β_{ij}. These are the parameters that are iterated in the minimization algorithm described in section 4. For purposes of interpretation it may be more convenient to consider *standardized parameters* $\tau_a^{(i)}$ and λ_{ij} defined by

$$\tau_a^{(i)} = \alpha_a^{(i)}/(1 + \sum_{j=1}^{k} \beta_{ij}^2)^{-\frac{1}{2}} \ , \tag{19}$$

$$\lambda_{ij} = \beta_{ij}/(1 + \sum_{j=1}^{k} \beta_{ij}^2)^{-\frac{1}{2}} \ . \tag{20}$$

The standardized and unstandardized parameters are in a one-to-one correspondence so one can obtain the unstandardized parameters from the standardized parameters by the formulas:

$$\alpha_a^{(i)} = \tau_a^{(i)}/(1 - \sum_{j=1}^{k} \lambda_{ij}^2)^{-\frac{1}{2}} \ , \tag{21}$$

$$\beta_{ij} = \lambda_{ij}/(1 - \sum_{j=1}^{k} \lambda_{ij}^2)^{-\frac{1}{2}} \ . \tag{22}$$

6 Model Test

If $F(\boldsymbol{\theta})$ has been minimized with respect to $\boldsymbol{\theta}$, one can use the likelihood ratio (LR) test statistic

$$\chi^2_{\mathrm{LR}} = 2\sum_r n_r \ln(p_r/\hat{\pi}_r) = 2N\sum_r p_r \ln(p_r/\hat{\pi}_r) = 2NF(\hat{\boldsymbol{\theta}}) \ , \tag{23}$$

to test the model, where $\hat{\boldsymbol{\theta}}$ is the estimated parameter vector and $\hat{\pi}_r = \pi_r(\hat{\boldsymbol{\theta}})$. Hence, this χ^2 is $2N$ times the minimum value of the fit function (13). If the model holds, this is distributed approximately as χ^2 with degrees of freedom equal to the number of different response patterns minus one minus the number of independent elements of $\boldsymbol{\theta}$.

Alternatively, one can use the goodness-of-fit (GF) test statistic

$$\chi^2_{\mathrm{GF}} = \sum_r [(n_r - N\hat{\pi}_r)^2/(N\hat{\pi}_r)] = N\sum_r (p_r - \hat{\pi}_r)^2/\hat{\pi}_r \ . \tag{24}$$

If the model holds, both statistics (23) and (24) have the same asymptotic distribution under H_0.

One can also evaluate the fit of the model to the univariate and bivariate marginals in the sample. Consider bivariate probabilities, let g and h be two pairs of variables and let a index the categories of variable g and b index the categories of variable h. Since the variables are conditionally independent for given $\boldsymbol{\xi}$, they are also pairwise conditionally independent. Thus, the conditional probability of $(x_g = a,\ x_h = b)$ is

$$\pi^{(gh)}_{ab}(\boldsymbol{\xi}) = [F(\alpha^{(g)}_a - \sum_{j=1}^{k}\beta_{gj}\xi_j) - F(\alpha^{(g)}_{a-1} - \sum_{j=1}^{k}\beta_{gj}\xi_j)][F(\alpha^{(h)}_b - \sum_{j=1}^{k}\beta_{hj}\xi_j) - F(\alpha^{(h)}_{b-1} - \sum_{j=1}^{k}\beta_{hj}\xi_j)] \tag{25}$$

The unconditional probability is

$$\pi^{(gh)}_{ab}(\boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \pi^{(gh)}_{ab}(\boldsymbol{\xi})h(\boldsymbol{\xi})d\boldsymbol{\xi} \tag{26}$$

Let $n^{(gh)}_{ab}$ be the number of cases with $x_g = a$ and $x_h = b$ and let $p^{(gh)}_{ab} = n^{(gh)}_{ab}/N$ be the corresponding proportion. To obtain a measure of bivariate fit for the pair of variables g and h one can compute the LR statistic

$$F^{(gh)}_{LR}(\hat{\boldsymbol{\theta}}) = 2N\sum_{a=1}^{m_g}\sum_{b=1}^{m_h} p^{(gh)}_{ab} \ln[p^{(gh)}_{ab}/\hat{\pi}^{(gh)}_{ab}] \tag{27}$$

The corresponding GF statistic is

$$F^{(gh)}_{GF}(\hat{\boldsymbol{\theta}}) = N\sum_{a=1}^{m_g}\sum_{b=1}^{m_h} (p^{(gh)}_{ab} - \hat{\pi}^{(gh)}_{ab})^2/\hat{\pi}^{(gh)}_{ab} \tag{28}$$

To get an overall measure of bivariate fit one can sum these statistics over all pairs of variables.

It should be pointed out that the bivariate statistics in (27) and (28) do not have asymptotic chi-square distributions because the bivariate likelihoods have not been maximized. Neither do their sums have asymptotic chi-square distributions. Nevertheless, the statistics are useful as measures of fit.

7 LISREL Implementation

7.1 Input

To perform factor analysis of ordinal variables, read the raw data, select the variables to be factor analyzed, and include the line

```
OFA [NF=k] [NOR/POM]
```

in the PRELIS command file.

Each of the options in brackets is optional.

- `NF=k` is used to specify the number of factors `k`. The default value is 1.
- One can specify either `NOR` or `POM`. If neither `NOR` nor `POM` is specified, `NOR` will be used.

This will be illustrated in the examples in section 8.

7.2 Output Files

Up to five output files are generated.

- The ordinary PRELIS output file is `INPUTFILE.OUT`. This contains the estimated unrotated standardized factor loadings, and various standardized rotated factor loadings (varimax, promax, reference variables solution). `INPUTFILE` is the name of the PRELIS command file in which the `OFA` command is included.
- An additional PRELIS output file `INPUTFILE.XXX` where `XXX` is `NOR`, or `POM` depending on which model is used. This contains various measures of univariate, bivariate, and multivariate measures of fit and estimates of the unstandardized and standardized solutions for both intercepts (thresholds) and factor loadings and their standard errors. It also contains some technical output (starting values and iteration progress).
- `BIVFITS.XXX` where `XXX` is `NOR` or `POM` containing detailed univariate and bivariate LR and GF fits.
- `MULFITS.XXX` where `XXX` is `NOR` or `POM` containing detailed LR and GF fits for all response patterns occuring in the sample.
- `INPUTFILE.FREQ` containing the data in frequency form. This is always obtained with ordinal variables in PRELIS (even without an `OFA` line), see Jöreskog (2002-2005)

This will be illustrated in the examples in section 8.

8 Examples

8.1 Example 1: Political Efficacy with Four Categories

In the first example we analyze the political efficacy variables described in section 1. The file **EFFICACY.RAW** in the **Ordinal** subfolder contains the data in text (ASCII) form. This can be analyzed using the following PRELIS command file (**EFFICACY1.PR2**).

```
Ordinal Factor Analysis of Six Efficacy Variables
DATA NI = 6 MI=8,9
LABELS
NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
RAWDATA=EFFICACY.RAW
CLABELS NOSAY - INTEREST 1=AS 2=A 3=D 4=DS 8=DK 9=NA
OFA POM NF=2
OUTPUT
```

The data values 8 and 9 represent *Don't Know* and *No Answer* responses, respectively. They are declared as missing values on the DATA line.

Alternatively, one can use the PRELIS system file EFFICACY.PSF and the following simpler command file (**EFFICACY2.PR2**).

```
Ordinal Factor Analysis of Six Efficacy Items
SY=EFFICACY.PSF
OFA NF=2 POM
OU
```

The output file **EFFICACY1.OUT** gives the following unrotated standardized factor loadings:

```
 Unrotated Factor Loadings

            Factor 1   Factor 2 Unique Var
            --------   -------- ----------
    NOSAY      0.876      0.000      0.233
   VOTING      0.699     -0.186      0.476
  COMPLEX      0.773      0.074      0.397
   NOCARE      0.868      0.381      0.102
    TOUCH      0.719      0.590      0.135
 INTEREST      0.755      0.571      0.103
```

Initially it was assumed that the factors $\xi_1, \ldots, \xi_k$ are uncorrelated and have variances 1. These assumptions can be relaxed somewhat and the factors may be correlated. If $k > 1$, the factor loadings in the matrix $\mathbf{B} = \beta_{ij}$ are not uniquely defined. Geometrically the factor loadings may be viewed as p points in a k-dimensional space. In this space the points are fixed but their coordinates can be referred to different factor axes. If the factor axes are orthogonal we say we have an *orthogonal solution*; if they are oblique we say that we have an *oblique solution* where the cosine of the angles between the factor axes are interpreted as correlations between the factors. In statistical terminology, an orthogonal solution corresponds to *uncorrelated factors* and an oblique solution corresponds to *correlated factors*.

To facilitate the interpretation of the factors one makes an orthogonal or oblique rotation of the factor axes. This rotation is usually guided by Thurstone's principle of simple structure which essentially states that only a small fraction of the loadings in each row and column of **B** should be large. Geometrically, this means that the factor axes pass through or near as many points as possible.

Let **T** be an arbitrary non-singular matrix of order $k \times k$ and let

$$\boldsymbol{\xi}^* = \mathbf{T}\boldsymbol{\xi} \qquad \mathbf{B}^* = \mathbf{B}\mathbf{T}^{-1} \, .$$

Thus, replacing $\boldsymbol{\xi}$ by $\boldsymbol{\xi}^*$ and $\mathbf{B}$ by $\mathbf{B}^*$ in (6) has no effect on the manifest variables. For the rotated factors to have unit variance, $\mathbf{T}$ must satisfy

$$diag(\mathbf{T}\mathbf{T}') = \mathbf{I} \,, \tag{29}$$

for an oblique solution and

$$\mathbf{T}\mathbf{T}' = \mathbf{I} \,, \tag{30}$$

for an orthogonal solution. For a review of several analytical rotation procedures, see Browne (2001).

PRELIS resolves this indeterminacy by fixing the loadings in the upper right corner of $\mathbf{B}$ to zero. In the example there is a fixed zero factor loading for NOSAY on Factor 2. If this is not done the information matrix will be singular and it will not be possible to obtain standard errors of the parameter estimates. (These standard errors are given in the output file **EFFICACY1.POM** described later.)

PRELIS provides three rotated solutions, one orthogonal (varimax) solution and two oblique solutions (promax and a reference variable rotation) which are given in the output file as

```
 Varimax-Rotated Factor Loadings

            Factor 1   Factor 2 Unique Var
            --------   -------- ----------
    NOSAY      0.752      0.448      0.233
   VOTING      0.696      0.199      0.476
  COMPLEX      0.626      0.459      0.397
   NOCARE      0.551      0.771      0.102
    TOUCH      0.315      0.875      0.135
 INTEREST      0.356      0.877      0.103

 Promax-Rotated Factor Loadings

            Factor 1   Factor 2 Unique Var
            --------   -------- ----------
    NOSAY      0.763      0.162      0.233
   VOTING      0.790     -0.111      0.476
  COMPLEX      0.602      0.238      0.397
   NOCARE      0.387      0.652      0.102
    TOUCH      0.054      0.894      0.135
 INTEREST      0.103      0.877      0.103

 Factor Correlations

            Factor 1   Factor 2
            --------   --------
 Factor 1      1.000
 Factor 2      0.642      1.000
```

```
 Reference Variables Factor Loadings

            Factor 1   Factor 2 Unique Var
            --------   -------- ----------
    NOSAY      0.685      0.277      0.233
   VOTING      0.724      0.000      0.476
  COMPLEX      0.534      0.332      0.397
   NOCARE      0.316      0.728      0.102
    TOUCH      0.000      0.930      0.135
 INTEREST      0.046      0.919      0.103

 Factor Correlations

            Factor 1   Factor 2
            --------   --------
 Factor 1      1.000
 Factor 2      0.584      1.000
```

The interpretation of these results is that Factor 1 represents *Internal Efficacy* and Factor 2 represents *External Efficacy*, with NOSAY, VOTING, and COMPLEX measuring Internal Efficacy and TOUCH and INTEREST measuring External Efficacy. NOCARE is a composite measure of both factors. The two oblique solutions suggest that the two factors correlate about 0.6. This is in line with the results in Jöreskog (2002-2005).

The results in the ***.OUT** file is intended for users who are mainly interested in the substantive results. For those interested in statistical and technical information about the performance of the models and the method, further results are obtained in the other output files.

Consider the output file **EFFICACY1.POM**. The following characteristics of the data and the model are reported first

```
 Number of First Order Frequencies     24

 Number of Second Order Frequencies    240

 Number of Zero Cells in Bivariate Tables =          2

 Bivariate Information Coverage Ratio =  0.992

 Sample Size =      1554.

 Number of Categories
   NOSAY:   4
  VOTING:   4
 COMPLEX:   4
  NOCARE:   4
   TOUCH:   4
INTEREST:   4
```

```
Total Number of Thresholds =   18
Number of Factors =    2
Number of Factor Loadings =     12
Number of Independent Factor Loadings =     11
Total Number of Independent Parameters =     29
```

Next the starting values for the FIML iterations are given:

```
Starting Values

Standardized Thresholds Tau^(i)_a
   NOSAY   -1.268   -0.235    1.420
  VOTING   -0.937    0.209    1.680
 COMPLEX   -0.844    0.726    1.754
  NOCARE   -1.023    0.160    1.788
   TOUCH   -0.971    0.530    2.119
INTEREST   -0.976    0.327    2.068

Standardized Factor Loadings Lambda_ij
   NOSAY    0.682    0.000
  VOTING    0.481   -0.155
 COMPLEX    0.548    0.044
  NOCARE    0.784    0.321
   TOUCH    0.616    0.520
INTEREST    0.662    0.530

Unstandardized Thresholds Alpha^(i)_a
   NOSAY   -1.733   -0.322    1.941
  VOTING   -1.086    0.242    1.947
 COMPLEX   -1.011    0.868    2.099
  NOCARE   -1.926    0.301    3.369
   TOUCH   -1.640    0.895    3.577
INTEREST   -1.840    0.617    3.899

Unstandardized Factor Loadings Beta_ij
   NOSAY    0.932    0.000
  VOTING    0.557   -0.180
 COMPLEX    0.656    0.053
  NOCARE    1.477    0.605
   TOUCH    1.039    0.878
INTEREST    1.247    0.999
```

These starting values are obtained as follows. The standardized thresholds are estimated from the univariate marginal distributions under underlying normality, see Jöreskog (2002-2005). The standardized factor loadings are estimated by factor analyzing the matrix of polychoric correlations by MINRES, see Jöreskog (2003). These estimates are consistent under the NOR model *but not* under the POM model. This is the reason why it usually takes a few more iterations to obtain POM estimates as compared to NOR estimates. The unstandardized estimates are obtained from the standardized estimates using the formulas (21) and (22).

The next lines in **EFFICACY1.POM** give the behavior of the iterations for FIML estimates. These lines are only of interest if something goes wrong. Details of the minimization algorithm are given in Jöreskog & Sörbom (1999, pp. 325–331). It happens occasionally that one or more of the thresholds goes to $-\infty$ or $+\infty$ during iterations. The iterations usually converge anyway after many iterations. This corresponds to Heywood cases in classical factor analysis in which a unique variance becomes zero. If this happens we recommend to reduce the number of factors.

The following lines give characteristics of the full information data, the minimum of the FIML fit function and the FIML LR and GF chi-square statistics. These are computed using the formulas in sections 3 and 6.

```
Number of Possible Response Patterns =     4096

Number of Distinct Response Patterns =      476

Full Information Coverage Ratio =  0.116

Minimum Fit Function Value  =     0.6087851832

-2ln L Under Model        =       18279.743
-2ln L Under Alternative  =       16387.638

LR Chi-square  with   446 Degrees of Freedom =          1892.10

GF Chi-square  with   446 Degrees of Freedom =        488392.42
```

These two chi-squares indicate that the overall fit is very bad. Note the extremely large GF statistic. The contributions of each response pattern to these chi-square statistics is given in the file **MULFITS.POM**. This file reveals that *one individual* with the response pattern

```
A DS DS DS AS AS
```

contributes 329485.12 to the GF chi-square. Obviously this is an extreme outlier.

It took 4.75 seconds to obtain the FIML estimates on a Pentium 4 computer with 1 GB memory and running at 3.0 Ghz. In **EFFICACY1.POM** the FIML estimates and standard error estimates of parameters are given in both unstandardized and standardized form as follows

```
FIML Estimates for Logistic Response Function (POM)

Unstandardized Thresholds Alpha^(i)_a
   NOSAY   -3.189   -0.557    3.587
  VOTING   -1.866    0.394    3.488
 COMPLEX   -1.743    1.516    3.801
  NOCARE   -3.514    0.613    6.111
   TOUCH   -2.983    1.640    6.523
INTEREST   -3.330    1.142    7.001
```

Unstandardized Factor Loadings Beta_ij

NOSAY	1.812	0.000
VOTING	1.013	-0.269
COMPLEX	1.226	0.117
NOCARE	2.723	1.194
TOUCH	1.953	1.603
INTEREST	2.348	1.777

Standard Errors for Unstandardized Thresholds Alpha^(i)_a

NOSAY	0.161	0.086	0.172
VOTING	0.088	0.066	0.155
COMPLEX	0.089	0.080	0.165
NOCARE	0.190	0.114	0.295
TOUCH	0.176	0.123	0.331
INTEREST	0.198	0.121	0.401

Standard Errors for Unstandardized Factor Loadings Beta_ij

NOSAY	0.134	0.000
VOTING	0.080	0.126
COMPLEX	0.088	0.114
NOCARE	0.186	0.186
TOUCH	0.146	0.197
INTEREST	0.167	0.228

Standardized Thresholds Tau^(i)_a

NOSAY	-1.540	-0.269	1.733
VOTING	-1.288	0.272	2.407
COMPLEX	-1.099	0.955	2.396
NOCARE	-1.120	0.195	1.948
TOUCH	-1.098	0.604	2.401
INTEREST	-1.071	0.367	2.251

Standardized Factor Loadings Lambda_ij

NOSAY	0.876	0.000
VOTING	0.699	-0.186
COMPLEX	0.773	0.074
NOCARE	0.868	0.381
TOUCH	0.719	0.590
INTEREST	0.755	0.571

Standard Errors for Standardized Thresholds Tau^(i)_a

NOSAY	0.065	0.040	0.078
VOTING	0.062	0.045	0.105
COMPLEX	0.056	0.054	0.118
NOCARE	0.058	0.035	0.102
TOUCH	0.046	0.039	0.103
INTEREST	0.049	0.037	0.102

```
Standard Errors for Standardized Factor Loadings Lambda_ij
   NOSAY    0.014    0.000
  VOTING    0.028    0.082
 COMPLEX    0.022    0.073
  NOCARE    0.030    0.058
   TOUCH    0.037    0.058
INTEREST    0.036    0.055
```

The univariate and bivariate LR and GF fit measures are given as follows in **EFFICACY1.POM**. These are computed from (27) and (28).

```
Univariate and Bivariate LR-Fits
   NOSAY    0.413
  VOTING  184.764    0.100
 COMPLEX   64.099   37.784    0.452
  NOCARE   54.190   73.144   51.350    1.541
   TOUCH   71.781   51.372   76.720   61.496    0.674
INTEREST   65.595   68.758   56.576   83.475   60.356    1.006

Total Univariate LR-Fit  =       4.186
Total  Bivariate LR-Fit  =    1061.461

Univariate and Bivariate GF-Fits
   NOSAY    0.412
  VOTING  248.050    0.100
 COMPLEX   92.201   40.930    0.447
  NOCARE   89.051   95.193   53.171    1.520
   TOUCH  119.897   62.140   99.830  133.086    0.651
INTEREST   94.250   85.354   63.741  254.082   81.219    0.971

Total Univariate GF-Fit  =       4.100
Total  Bivariate GF-Fit  =    1612.195
```

These fit measures indicate that the fit is very bad, particularly for the pair NOSAY and VOTING. Detailed information about the fit of each cell in the univariate and bivariate contingency table for each variable and each pair of variables is given in the file **BIVFITS.POM**.

8.2 Example 2: Political Efficacy with Two Categories

In view of the fact that many response patterns have very small frequency of occurrence, it may be a good idea to dichotomize the variables. In the PRELIS command file (**EFFICACY3.PR2**) the two categories *Agree Strongly* and *Agree* are collapsed into one *Agree* category and the two categories *Disagree* and *Disagree Strongly* are collapsed into one *Disagree* category and a factor analysis of these dichotomized variables are specified at the same time.

```
Ordinal Factor Analysis of Six Dichotomized Efficacy Variables
DATA NI = 6 MI=8,9
LABELS
```

```
NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
RAWDATA=EFFICACY.RAW
RE NOSAY-INTEREST OLD=1,2,3,4 NEW=1,1,2,2
CLABELS NOSAY - INTEREST 1=A 2=D
OFA NF=2 POM
OUTPUT
```

The factor loadings in the the output file **EFFICACY3.OUT** are as follows

```
Unrotated Factor Loadings

           Factor 1   Factor 2 Unique Var
           --------   -------- ----------
    NOSAY     0.923      0.000      0.149
   VOTING     0.804     -0.232      0.300
  COMPLEX     0.642      0.125      0.572
   NOCARE     0.847      0.438      0.091
    TOUCH     0.765      0.509      0.155
 INTEREST     0.713      0.629      0.096

Varimax-Rotated Factor Loadings

           Factor 1   Factor 2 Unique Var
           --------   -------- ----------
    NOSAY     0.784      0.487      0.149
   VOTING     0.805      0.227      0.300
  COMPLEX     0.480      0.445      0.572
   NOCARE     0.488      0.819      0.091
    TOUCH     0.381      0.836      0.155
 INTEREST     0.274      0.910      0.096

Promax-Rotated Factor Loadings

           Factor 1   Factor 2 Unique Var
           --------   -------- ----------
    NOSAY     0.767      0.216      0.149
   VOTING     0.907     -0.114      0.300
  COMPLEX     0.405      0.313      0.572
   NOCARE     0.252      0.769      0.091
    TOUCH     0.111      0.842      0.155
 INTEREST    -0.055      0.986      0.096

Factor Correlations

           Factor 1   Factor 2
           --------   --------
 Factor 1     1.000
 Factor 2     0.655      1.000
```

```
Reference Variables Factor Loadings

          Factor 1   Factor 2 Unique Var
          --------   -------- ----------
   NOSAY     0.723      0.303      0.149
  VOTING     0.836      0.000      0.300
 COMPLEX     0.393      0.353      0.572
  NOCARE     0.274      0.778      0.091
   TOUCH     0.147      0.832      0.155
INTEREST     0.000      0.951      0.096

Factor Correlations

          Factor 1   Factor 2
          --------   --------
Factor 1     1.000
Factor 2     0.538      1.000
```

These results have roughly the same interpretation as those in **EFFICACY1.OUT**.

As shown in the file **EFFICACY3.POM** it took only 0.14 seconds to obtain the POM-FIML solution and the fit of the model is now much better:

```
Number of Possible Response Patterns =       64

Number of Distinct Response Patterns =       62

Full Information Coverage Ratio =  0.969

Minimum Fit Function Value  =     0.0203242397

-2ln L Under Model        =        10369.974
-2ln L Under Alternative  =        10306.806

LR Chi-square  with    44 Degrees of Freedom =            63.17

GF Chi-square  with    44 Degrees of Freedom =            58.12
```

Note that all but two of the response patterns occur in the sample. The file **EFFICACY3.FREQ** reveals that 46 out of the 62 response patterns have a frequency of occurrence of 5 or more.

Based on the 62 response patterns we get a LR chi-square of 63.17 with 44 degrees of freedom. The corresponding GF chi-square is 58.12. Both chi-squares indicate a reasonably good fit.

The fit to the univariate and bivariate marginals is also quite good.

```
Univariate and Bivariate LR-Fits
   NOSAY    0.027
  VOTING    0.063    0.001
 COMPLEX    0.535    0.062    0.004
  NOCARE    0.380    1.323    1.121    0.045
   TOUCH    0.197    0.164    1.391    0.287    0.001
INTEREST    0.403    0.083    0.072    0.144    0.134    0.013

Total Univariate LR-Fit  =        0.091
Total  Bivariate LR-Fit  =        6.359

Univariate and Bivariate GF-Fits
   NOSAY    0.027
  VOTING    0.063    0.001
 COMPLEX    0.538    0.062    0.004
  NOCARE    0.378    1.331    1.107    0.045
   TOUCH    0.198    0.163    1.389    0.288    0.001
INTEREST    0.405    0.083    0.072    0.144    0.133    0.013

Total Univariate GF-Fit  =        0.090
Total  Bivariate GF-Fit  =        6.354
```

Note that the LR and GF statistics are quite close. This is typically the case when the model fits well.

References

Aish, A.M., & Jöreskog, K.G. (1990). A panel model for political efficacy and responsiveness: An application of LISREL 7 with weighted least squares. *Quality and Quantity*, **24**, 405-426.

Bartholomew, D., & Knott, M. (1999). *Latent Variable Models and Factor Analysis*. Second edition. Kendall Library of Statistics. London: Arnold.

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

Browne, M.W. (2001) An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, **36**, 111–150.

Jöreskog, K. G. (2002-2005). Structural equation modeling with ordinal variables using LISREL. Available at *www.ssicentral.com/lisrel/techdocs.ordinal.pdf*.

Jöreskog, K. G. (2003). Factor Analysis by MINRES. Available at *www.ssicentral.com/lisrel/techdocs.minres.pdf*.

Jöreskog, K. G. (2006). *Factor analysis and its extensions*. In R. Cudeck & R. MacCallum (Eds.) *Factor Analysis at 100*. Lawrence Erlbaum Publishers, forthcoming.

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, **36**, 347–387.

Jöreskog, K.G., & Sörbom, D. (1999). LISREL 8 *User's Reference Guide*. Second Edition. Lincolnwood: Scientific Software International.

Jöreskog, K.G., Sörbom, D., Du Toit, S., & Du Toit, M. (2003). LISREL 8*: New Statistical Features*. Third printing with revisions. Lincolnwood: Scientific Software International.

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

Moustaki, I. (2006). Factor analysis and latent structure of categorical and metric data. In R. Cudeck & R. MacCallum (Eds.) *Factor Analysis at 100*. Lawrence Erlbaum Publishers, forthcoming.

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

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, **14**, 59–71.

Muraki, E., & Carlson, E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, **19**, 73–90.