Factor Analysis of Ordinal Variables: A Comparison of Three Approaches

Article *in* Multivariate Behavioral Research · February 2001

DOI: 10.1207/S15327906347-387

2 authors:

Karl G. Jöreskog
Uppsala University

Ir Moustaki
The London School of Economics and Political Science

Multivariate Behavioral Research, 36 (3), 347-387

Factor Analysis of Ordinal Variables: A Comparison of Three Approaches

Karl G. Jöreskog
Department of Information Science
Uppsala University

Irini Moustaki
Department of Statistics
London School of Economics

Theory and methodology for exploratory factor analysis have been well developed for continuous variables. In practice, observed or measured 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. In this article we describe four approaches to factor analysis of ordinal variables which take proper account of ordinality and compare three of them with respect to parameter estimates and fit. The comparison is made both in terms of their relative methodological advantages and in terms of an empirical data example and two generated data examples. In particular, we discuss the issue of how to test the model and to measure model fit.

The basic idea of factor analysis is the following. For a given set of response variables x_1, ..., x_p one wants to find a set of latent factors z_1, ..., z_k, fewer in number than the observed variables, that contain essentially the same information. The latent factors are supposed to account for the dependencies among the response variables in the sense that if the factors are held fixed, the observed variables would be independent. If both the response variables and the latent factors are normally distributed with zero means and unit variances, this leads to the model (see Jöreskog, 1979)

$$E\left(x_i | z_1, z_2, \ldots, z_k\right) = \lambda_{i1} z_1 + \lambda_{i2} z_2 + \ldots + \lambda_{ik} z_k, \tag{1}$$

$$E\left(x_i x_j | z_1, z_2, \ldots, z_k\right) = 0, \quad i \neq j. \tag{2}$$

Correspondence concerning this article should be addressed to Karl G. Jöreskog, Department of Information Science, Uppsala University, P.O. Box 513, SE-75120 Uppsala, Sweden. Electronic mail may be sent to Karl.Joreskog@dis.uu.se.

If the factors are independent, it follows that the correlation ρ_{ij} between x_i and x_j is

$$(3)\qquad \rho_{ij} = \sum_{l=1}^{k} \lambda_{il}\lambda_{jl}.$$

Equation 1 is a suitable representation of the factor analysis model if the response variables x_i are continuous variables measured on an interval or ratio scale. It cannot be used, however, if the response variables x_i are ordinal or nominal. In those cases, one must instead specify the probability of each response pattern as a function of z_1, z_2, ..., z_k:

$$(4)\qquad \Pr\left(x_1 = a_1, x_2 = a_2, \ldots, x_p = a_p | z_1, z_2, \ldots, z_k\right) = f\left(z_1, z_2, \ldots, z_k\right),$$

where a_1, a_2, ..., a_p represent the different response categories of x_1, x_2, ..., x_p, respectively. In this article we consider the case of ordinal response variables and continuous latent factors.

There are two main approaches for analyzing ordinal variables using latent variable models and within these there are different variants.

The *underlying response variable approach* assumes that each observed ordinal variable is generated by an underlying unobserved continuous variable assumed to be normally distributed. The contributors to this approach discuss the estimation of the model either in two or three stages. Muthen (1984) proposed a three-stage estimation method for handling mixed type variables (including ordinal polytomous). At the first stage, first order statistics such as thresholds, means and variances are estimated by maximum likelihood. In the second stage, second order statistics such as polychoric correlations are estimated by conditional maximum likelihood for given first stage estimates. At the third stage the parameters of the structural part of the model are estimated using a generalized least squares method. Jöreskog (1990,1994) also proposed a three stage estimation method. At the third stage of this method the parameters of the factor model are estimated by a weighted least squares method where the weight matrix is an estimate of the inverse of the asymptotic covariance matrix of the polychoric correlations. As an extension of this work Lee, Poon, and Bentler (1990, 1992) proposed a two-stage estimation procedure. Arminger and Küsters (1988) have also adopted an underlying response variable approach.

The *the response function approach* specifies the conditional distribution of the complete p-dimensional response pattern as a function of the latent factors and makes the assumption that responses to different variables are independent for given latent factors (conditional independence). This approach is particularly developed within item response theory for dichotomous variables and a single latent factor. The response function is either the logit or the probit. Samejima (1969) discusses a logit and a probit model that can be used to model ordinal responses. This unidimensional graded response model was a starting point for research that followed. The response functions proposed in that article are a probit model and a logistic model with a slope parameter (discrimination parameter) for each item and an item response parameter (difficulty parameter). Muraki (1990) discusses the estimation of the unidimensional graded response model using a marginal maximum likelihood method with the EM method. In his parameterization of the model the item response parameter is written as two parameters: the item location parameter, and the category threshold parameters and this model is called a rating-scale model. This work was further extended in a article by Muraki and Carlson (1995) where a multidimensional probit model for ordinal variables is discussed.

The four approaches that are reviewed in this article are: (a) UMN, "The Underlying Multivariate Normal Approach", (b) UBN, "The Underlying Bivariate Normal Approach", (c) NOR, "The Normal Ogive Approach", and (d) POM, "The Proportional Odds Model Approach".

UMN is the full information maximum likelihood (FIML) approach of Lee, Poon, and Bentler (1990) here applied to the factor analysis model. This involves the numerical evaluation of p-dimensional integrals which is not computationally feasible for general factor analysis problems.

Since UMN is not feasible we propose a limited information method for estimating the same model, namely UBN which maximizes the sum of all univariate and bivariate marginal likelihoods. It is of interest to compare this approach to the two FIML methods NOR and POM.

NOR and POM are similar in the sense that they assume conditional independence. They differ only in terms of a different cumulative response function which is the normal for NOR and the logistic for POM.

POM is based on a logistic regression model and uses a full information method for estimation. Moustaki (2000) presents a general class of models of which POM is a special case. POM may also be seen as an extension of the class of models considered by Moustaki and Knott (2000).

The three approaches, UBN, NOR, and POM, attempts to estimate all parameters in one step. NOR and POM uses all the data, whereas UBN uses only the data in the univariate and bivariate margins.

UBN, NOR, and POM are not available in any computer program for structural equation modeling. The unidimensional POM model can be fitted with MULTILOG (Thissen, 1991). The results reported in this article have been obtained with our own FORTRAN program and, as will be explained, we find that it is not computationally feasible to apply the UMN method to more than four variables.

General Framework

Let $x_1, x_2, ..., x_p$ denote the observed ordinal variables and let m_i denote the number of response categories of variable i. We write $x_i = a$ to mean that x_i belongs to the ordered category a, $a = 1, 2, ..., m_i$. The actual score values in the data may be arbitrary and are irrelevant as long as the ordinal information is retained. That is, low scores correspond to low-order categories of x_i and high scores correspond to high-order categories.

There are

$$\prod_{i=1}^{p} m_i$$

possible response patterns. Let $\mathbf{x}_r = (x_1 = a_1, x_2 = a_2, ..., x_p = a_p)$ represent anyone of these. The model specifies the probability $\pi_r = \pi_r(\boldsymbol{\theta}) > 0$ of $\mathbf{x}_r$ as a function of the independent parameters $\boldsymbol{\theta}$, for all possible response patterns r. These probabilities are subject to $\sum \pi_r = 1$, for all $\boldsymbol{\theta}$. The different approaches to be considered differ in the way $\pi_r(\boldsymbol{\theta})$ is specified and in the way the model is estimated.

Let n_r be the observed sample frequency of response pattern r. Then the log likelihood may be written as

$$\ln L = \sum_r n_r \ln \pi_r = N \sum_r p_r \ln \pi_r \tag{5}$$

where $N = \sum_r n_r$ is the sample size and $p_r = n_r/N$ is the sample proportion of response pattern r.

If there is no model so that the π_r are unconstrained, the maximum of $\ln L$ is

$$\ln L_1 = \sum_r n_r \ln p_r = \sum_r p_r \ln p_r .$$

Instead of maximizing Equation 5, it is convenient to minimize the fit function

$$F(\boldsymbol{\theta}) = \sum_r p_r [\ln p_r - \ln \pi_r(\boldsymbol{\theta})] = \sum_r p_r \ln[p_r / \pi_r(\boldsymbol{\theta})]. \tag{6}$$

Obviously, the sum in Equation 6 need only be over all response patterns present in the data.

To minimize Equation 6 the gradient vector and the information matrix are needed. The gradient vector is

$$\partial F(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} = -\sum_r \left[p_r / \pi_r(\boldsymbol{\theta}) \right] \partial \pi_r(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}, \tag{7}$$

and the Hessian matrix is

$$\partial^2 F / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' = \sum_r \left\{ -\left[p_r / \pi_r(\boldsymbol{\theta}) \right] \partial^2 \pi_r(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' + \left[p_r / \pi_r^2(\boldsymbol{\theta}) \right] \partial \pi_r(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \pi_r(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}' \right\}. \tag{8}$$

Taking *plim* of the Hessian, and noting that under the model, $plim\, p_r = \pi_r(\boldsymbol{\theta})$, $\sum \partial \pi_r(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} = 0$ and $\sum \partial^2 \pi_r(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' = 0$, one obtains the information matrix

$$plim\, \partial^2 F / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' = \sum \left[1 / \pi_r(\boldsymbol{\theta}) \right] \partial \pi_r(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \pi_r(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'. \tag{9}$$

Let $\boldsymbol{\theta}^{(0)}$ represent 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

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

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

$$\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)} - \alpha^{(s)} \mathbf{E}^{(s)} \mathbf{g}^{(s)}. \tag{10}$$

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

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

$$\chi^2_{\mathrm{LR}} = 2 \sum_r n_r \ln(p_r / \hat{\pi}_r) = 2N \sum_r p_r \ln(p_r / \hat{\pi}_r) = 2NF(\hat{\boldsymbol{\theta}}), \tag{11}$$

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 just $2N$ times the minimum value of the fit function in Equation 6. 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

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

If the model holds, both statistics in Equations 11 and 12 have the same asymptotic distribution under H_0. As will be argued in the section entitled "Testing the Model," there are certain advantages of using the LR statistic rather than the GF statistic.

The UMN Approach

The UMN approach employs the classical factor analysis model:

(13) $$x_i^* = \lambda_{i1} z_1 + \lambda_{i2} z_2 + \cdots + \lambda_{ik} z_k + u_i , \quad i = 1,2,\ldots,p,$$

where u_i is an error term representing a specific factor and measurement error and x_i^* is an unobserved continuous response variable *underlying* the ordinal variable x_i. In classical factor analysis x^*_i is assumed to be directly observed but here it is unobserved.

The connection between the ordinal variable x_i and the underlying variable x_i^* is

(14) $$x_i = a \Leftrightarrow \tau^{(i)}_{a-1} < x_i^* \le \tau^{(i)}_a , a = 1,2,\ldots,m_i ,$$

where

$$\tau^{(i)}_0 = -\infty, \tau^{(i)}_1 < \tau^{(i)}_2 < \ldots < \tau^{(i)}_{m_i-1}, \ \tau^{(i)}_{m_i} = +\infty,$$

are parameters called threshold values. For variable x_i with m_i categories, there are m_i - 1 threshold parameters.

Since only ordinal information is available about x_i^*, the mean and variance of x_i^* are not identified and are therefore set to zero and one, respectively.[1] It is further assumed that $z_1, \ldots, z_k, u_1, \ldots, u_p$ are independent and normally distributed with $z_j \sim N(0,1)$ and $u_i \sim N(0,\psi_i^2)$. Since x_i^* is standardized it follows that $\psi_i^2 = 1 - \sum_{j=1}^{k} \lambda_{ij}^2$. Furthermore, it follows that $x_1^*, \ldots, x_p^*$ has a multivariate normal distribution (hence the name UMN) with zero means, unit variances and correlation matrix $\mathbf{P} = (\rho_{ij})$, where $\rho_{ij} = \sum_{l=1}^{k} \lambda_{il}\lambda_{jl}$.

[1] Other parameterizations are possible, see for example, Lee, Poon, and Bentler (1990). See also the section entitled "The NOR Approach."

The parameters of the model are the thresholds $\tau_a^{(i)}$, $i = 1, 2, \ldots, p$, $a = 1, 2, \ldots, m_i - 1$ and the factor loadings λ_{ij}, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, k$. Since, for $k > 1$, the matrix $\boldsymbol{\Lambda} = (\lambda_{ij})$ of order $p \times k$ is determined only up to an orthogonal transformation of order $k \times k$, there are only $pk - k(k - 1)/2$ independent factor loadings. Hence, the total number of independent parameters of the model is $\sum_{i=1}^{p} m_i - p + pk - k(k-1)/2$.

From Equation 14 and the multivariate normality of $x_1^*, \ldots, x_p^*$, it follows that the probability $\pi_r(\boldsymbol{\theta})$ of a general p-dimensional response pattern is

$$\pi_r(\boldsymbol{\theta}) = \Pr(x_1 = a_1, x_2 = a_2, \ldots, x_p = a_p) \tag{15}$$

$$= \int_{\tau_{a_1-1}^{(1)}}^{\tau_{a_1}^{(1)}} \int_{\tau_{a_2-1}^{(2)}}^{\tau_{a_2}^{(2)}} \cdots \int_{\tau_{a_p-1}^{(p)}}^{\tau_{a_p}^{(p)}} \phi_p(u_1, u_2, \ldots, u_p|\mathbf{P}) du_1, du_2, \ldots, du_p \ , \tag{16}$$

where the integral Equation 16 is over the p -dimensional normal density function $\phi_p(u_1, u_2, \ldots, u_p|\mathbf{P})$ with zero means, unit variances and correlation matrix $\mathbf{P}$, and the integration limits of the i^{th} dimension is from $\tau_{a_i-1}^{(i)}$ to $\tau_{a_i}^{(i)}$.

Lee, Poon, and Bentler (1990, eq. 2) give an expression for $\pi_r(\boldsymbol{\theta})$ as a linear combination of 2^p terms, each term being a value of the multinormal distribution function. To use this expression, however, would make the computational burden much heavier since it takes about the same time to compute the multinormal distribution function as it takes to compute the integral in Equation 16 directly.

To obtain full information maximum likelihood estimates in the UMN approach we minimize the fit function Equation 6 with respect to the parameters $\boldsymbol{\theta}$ using Equation 16. This requires the evaluation of the p-dimensional integral in Equation 16 for each response pattern in the sample at several points of the parameter space. For $p \leq 7$ one can use a double precision version of the algorithm MULNOR by Schervish (1984) to evaluate these integrals. The time for these integrations increases rapidly with p, so that the UMN is not computationally feasible with $p > 4$.

The UBN Approach

The UBN approach makes the same assumptions as the UMN approach[2] but uses only the data in the univariate and bivariate margins to estimate the model.

From Equation 14 and the multivariate normality of $x_1^*, \ldots, x_p^*$, it follows immediately that the probability $\pi_a^{(g)}$ of a response in category a on variable g is

[2] Actually, it is only necessary to assume that all pairs of underlying response variables have a bivariate normal distribution.

$$\pi_a^{(g)}(\boldsymbol{\theta}) = \int_{\tau_{a-1}^{(g)}}^{\tau_a^{(g)}} \phi(u)du, \tag{17}$$

where $\phi(u)$ is the standard normal density function, and that the probability $\pi_{ab}^{(gh)}$ of a response in category a on variable g and a response in category b on variable h is

$$\pi_{ab}^{(gh)}(\boldsymbol{\theta}) = \int_{\tau_{a-1}^{(g)}}^{\tau_a^{(g)}} \int_{\tau_{b-1}^{(h)}}^{\tau_b^{(h)}} \phi_2(u, \upsilon | \rho_{gh}) du d\upsilon, \tag{18}$$

where $\phi_2(u,\upsilon|\rho)$ is the density function of the standardized bivariate normal distribution with correlation ρ.

The UBN approach minimizes the sum of all univariate and bivariate fit functions (equivalent to maximizing the sum of all univariate and bivariate log-likelihoods):

$$F_{UBN}(\boldsymbol{\theta}) = \sum_{g=1}^{p} \sum_{a=1}^{m_g} p_a^{(g)} \ln\left[p_a^{(g)} / \pi_a^{(g)}(\boldsymbol{\theta})\right] + \sum_{g=2}^{p} \sum_{h=1}^{g-1} \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} p_{ab}^{(gh)} \ln\left[p_{ab}^{(gh)} / \pi_{ab}^{(gh)}(\boldsymbol{\theta})\right], \tag{19}$$

where $p_a^{(g)}$ is the sample proportion of category a in the univariate margin of variable g, $p_{ab}^{(gh)}$ is the sample proportion of category a and b in the bivariate margin of variables g and h, and $\pi_a^{(g)}(\boldsymbol{\theta})$ and $\pi_{ab}^{(gh)}(\boldsymbol{\theta})$ are given by Equations 17 and 18, respectively. The integrals in Equations 17 and 18 can be computed directly from the univariate and bivariate distribution functions. Only data in the univariate and bivariate margins are used. This approach is quite feasible in that it can handle a large number of variables as well as a large number of factors.

Other approaches which also use only first and second order moments is that used in PRELIS/LISREL and described by Jöreskog (1990, 1994) and that used in MPLUS and described by Muthen (1984) and Muthen and Satorra (1995) in a more general setting. These are three-step procedures based on underlying normally distributed variables. In the first step, the thresholds are estimated from the univariate margins of the observed variables. In the second step, the polychoric correlations are estimated from the bivariate margins of the observed variables for given thresholds. In the third step, the factor model is estimated from the polychoric correlations by weighted least squares using a weight matrix which is the inverse of an estimate of the asymptotic covariance matrix of the polychoric correlations. The asymptotic covariance matrix is often unstable in small samples, particularly if there are zero or small frequencies in the bivariate margins. By contrast, the UBN approach estimates the thresholds and the factor loadings in one single step from the univariate and bivariate margins without the use of a weight matrix.

The NOR Approach

The NOR approach specifies that the conditional probability of a response in category s or lower on variable i is

$$\gamma_s^{(i)}(\mathbf{z}) = \int_{-\infty}^{\left[\alpha_s^{(i)} - \beta_{i1} z_1 - \beta_{i2} z_2 - \cdots - \beta_{ik} z_k\right]} \phi(u)du \tag{20}$$

$$= \Phi\left[\alpha_s^{(i)} - \sum_{j=1}^{k} \beta_{ij} z_j\right], \tag{21}$$

where $\Phi(u)$ is the standard normal distribution function.

The $\alpha_s^{(i)}$ are intercept parameters. There is one such intercept parameter for each variable and each category. To define ordinality properly, the intercept parameters must satisfy

$$\alpha_1^{(i)} < \alpha_2^{(i)} \ldots < \alpha_{m_{i-1}}^{(i)} < \alpha_{m_i}^{(i)} = \infty.$$

The β_{ij} parameters are the factor loadings. The number of independent parameters is $\Sigma_{i=1}^{p} m_i$ - p + pk - $k(k\text{-}1)/2$, which is the same as in the UMV and UBN approaches. As will be explained later, the parameters $\alpha_s^{(i)}$ and β_{ij} are unstandardized parameters.

In the same way we obtain the conditional probability $\pi_s^{(i)}(\mathbf{z})$ for category a of variable i:

$$\pi_a^{(i)}(\mathbf{z}) = \Phi\left[\alpha_a^{(i)} - \sum_{j=1}^{k} \beta_{ij} z_j\right] - \Phi\left[\alpha_{a-1}^{(i)} - \sum_{j=1}^{k} \beta_{ij} z_j\right] \tag{22}$$

It is convenient to refer to Equation 21 as the *cumulative response function* and to Equation 22 as the *category response function*.

For the case of $k = 1$, Figure 1 shows four typical cumulative response functions $\gamma_s^{(i)}(z)$ corresponding to the parameter values given in Table 17. The corresponding category response functions $\pi_\alpha^{(i)}(z)$ are shown in Figure 2. Each $\gamma_s^{(i)}(z)$ goes to 1 as z goes to -∞ and to 0 as z goes to +∞. The tendency to go to -∞ and +∞ is faster for larger β. Each $\pi_\alpha^{(i)}(z)$ looks like a normal density function. For different values of a with the same β these curves are shifted vertically. The curves are steeper for larger β.

The connection between the UMN and NOR approaches are as follows. From Equation 13 and the UMN assumptions, it follows that, conditional on

z, $x_i^* \sim N(\lambda_{i1}z_1 + \lambda_{i2}z_2 + \ldots + \lambda_{ik}z_k, \psi_i^2)$. Let $\alpha_a^{(i)} = \tau_a^{(i)}/\psi_i$ and $\beta_{ij} = \lambda_{ij}/\psi_i$. We shall refer to $\tau_a^{(i)}$ and λ_{ij} as *standardized parameters* and to $\alpha_a^{(i)}$ and β_{ij} as *unstandardized parameters*. In terms of the unstandardized parameters, the variance of x_i^* is $1 + \Sigma_{j=1}^k \beta_{ij}^2$. Hence, we can obtain the standardized parameters from the unstandardized parameters using the relationships

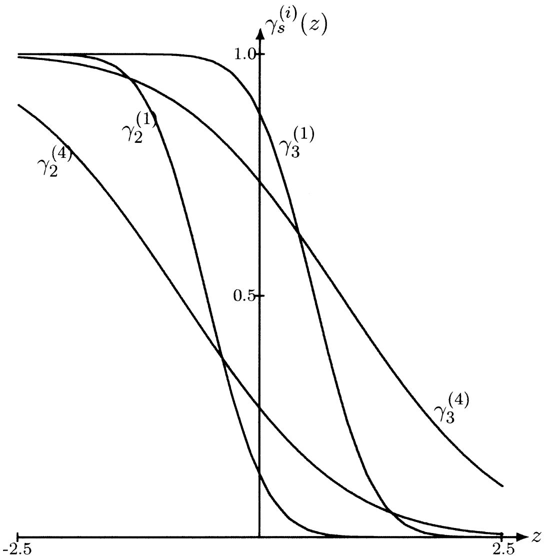

Figure 1
NOR: Four Cumulative Response Functions $\gamma_s^{(i)}z$

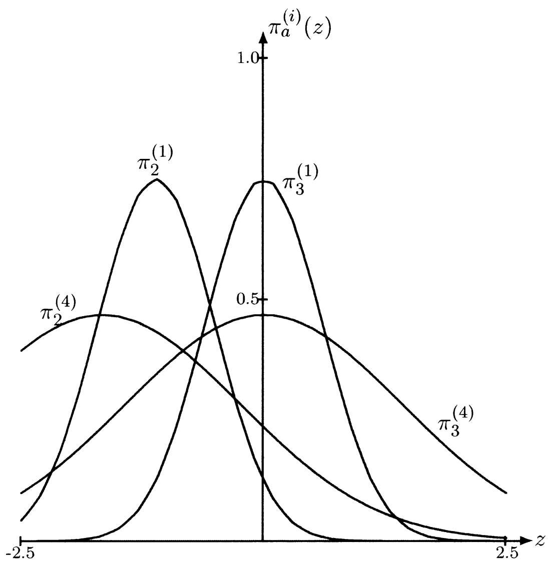

Figure 2
NOR: Four Category Response Functions $\pi_a^{(i)}z$

(23) $$\tau_a^{(i)} = \alpha_a^{(i)}\left(1+\sum_{j=1}^{k}\beta_{ij}^2\right)^{-\frac{1}{2}},$$

(24) $$\lambda_{ij} = \beta_{ij}\left(1+\sum_{j=1}^{k}\beta_{ij}^2\right)^{-\frac{1}{2}},$$

Similarly, one can obtain the unstandardized parameters from the standardized ones by

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

$$\beta_{ij} = \lambda_{ij}\left(1-\sum_{j=1}^{k}\lambda_{ij}^2\right)^{-\frac{1}{2}}.$$

These relationships were noted by Takane and de Leeuw (1987) and are given in Bartholomew and Knott (1999).

If the NOR model holds, then, conditional on $\mathbf{z}$, $\mathbf{x}^* \sim N(\mathbf{\Lambda z}, \mathbf{\Psi}^2)$ where $\mathbf{\Psi}$ is a diagonal matrix with diagonal elements $\psi_1, \psi_2, \ldots, \psi_p$. Integrating this over $\mathbf{z}$ gives $\mathbf{x}^* \sim N(\mathbf{0}, \mathbf{\Lambda\Lambda}' + \mathbf{\Psi}^2) = N(\mathbf{0},\mathbf{P})$, which shows that $\mathbf{x}^*$ has the distribution assumed in the UMV model.

Let $\mathbf{x}_r = (x_1 = a_1, x_2 = a_2, \ldots, x_p = a_p)$ represent a full response pattern. Under the assumption of conditional independence, the conditional probability, for given $\mathbf{z}$, of the response pattern $\mathbf{x}_r$ is

(25) $$\pi_r(\mathbf{z}) = \prod_{i=1}^{p}\pi_{a_i}^{(i)}(\mathbf{z}) = \prod_{i=1}^{p}\left[\gamma_{a_i}^{(i)}(\mathbf{z}) - \gamma_{a_i-1}^{(i)}(\mathbf{z})\right].$$

The unconditional probability π_r of the response pattern $\mathbf{x}_r$ is obtained by integrating $\pi_r(\mathbf{z})$ over the k-dimensional factor space:

(26) $$\pi_r(\theta) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\pi_r(\mathbf{z})h(\mathbf{z})d\mathbf{z},$$

where $h(\mathbf{z})$ is the density function of $\mathbf{z}$. The latent variables are assumed to be independent with standard normal distributions, so that $h(\mathbf{z}) = \Pi_{j=1}^{k}\phi(z_j)$.

The integral in Equation 26 can be approximated to any practical degree of accuracy by Gauss-Hermite quadrature, that is,

$$\pi_r = \sum_{t_1=1}^{\nu_1} \cdots \sum_{t_k=1}^{\nu_k} \pi_r(z_{1t_1}, \cdots, z_{kt_k}) h(z_{1t_1}) \cdots h(z_{kt_k}), \tag{27}$$

where $z_{1t_1}, \ldots, z_{kt_k}$ are tabled quadrature nodes, $h(z_{1t_1}), \ldots, h(z_{kt_k})$ are the corresponding weights and ν_1, ν_2, ..., ν_k are the number of quadrature points along each factor dimension.

For one factor, we use 48 quadrature points and for two factors we use 8 points along each dimension. For more than two factors one can use the general routine HRMSYM by Genz and Keister (1996).[3]

To obtain maximum likelihood estimates under the NOR model, the fit function in Equation 6 is minimized using the Fisher scoring algorithm described in the section entitled "General Framework".

The POM Approach

In this approach there are no underlying response variables. The model described here comes from item response theory (IRT). A variety of IRT models have been proposed in the literature for handling ordinal variables within the framework of latent factor models. The work that comes closest to ours is that by Muraki and Carlson (1995) who discusses a multi-dimensional IRT model for polytomous items based on Samejima's (1969) graded response model and uses the normal ogive. Our model for ordinal variables is based on the proportional odds model discussed in McCullagh (1980) within the framework of regression analysis. Moustaki (2000) discusses the POM model within the framework of exponential family distributions.

As before, let x_1, x_2, ..., x_p denote the observed ordinal variables and let m_i denote the number of response categories of variable i. The m_i ordered categories of variable i have probabilities $\pi_1^{(i)}(\mathbf{z})$, $\pi_2^{(i)}(\mathbf{z})$, ..., $\pi_{m_i}^{(i)}(\mathbf{z})$ which are functions of $\mathbf{z}$, the vector of the latent factors of order $k \times 1$.

The *proportional odds model* (POM) is the same as the logistic model:

$$\ln\left[\frac{\gamma_s^{(i)}(\mathbf{z})}{1-\gamma_s^{(i)}(\mathbf{z})}\right] = \alpha_s^{(i)} - \sum_{j=1}^{k} \beta_{ij} z_j, s=1,\ldots,m_i-1, \tag{28}$$

[3] Available at www.sci.wsu.edu/math/faculty/genz.

where, as before, $\gamma_s^{(i)}(\mathbf{z}) = Pr(x_i \leq s) = \pi_1^{(i)}(\mathbf{z}) + \ldots + \pi_s^{(i)}(\mathbf{z})$ is the probability of a response in category s or lower on variable i. The name proportional odds model comes from the fact that, in the one-factor case, the difference between two cumulative logits, that is, the left side of Equation 28, for two persons with factor scores z_1 and z_2 is proportional to $z_1 - z_2$. Note that there is one intercept parameter $\alpha_s^{(i)}$ for each category whereas the slope parameters β_{ij} remain the same across categories of the same variable. The π's are obtained from the γ's by the relationships $\pi_1^{(i)}(\mathbf{z}) = \gamma_1^{(i)}(\mathbf{z})$ and $\pi_a^{(i)}(\mathbf{z}) = \gamma_a^{(i)}(\mathbf{z}) - \gamma_{a-1}^{(i)}(\mathbf{z})$, $a = 2, 3, \ldots, m_i$.

As before, the intercept parameters $\alpha_s^{(i)}$ must satisfy

$$\alpha_1^{(i)} < \alpha_2^{(i)} \ldots < \alpha_{m_{i-1}}^{(i)} < \alpha_{m_i}^{(i)} = \infty.$$

The β_{ij} parameters are the factor loadings. The number of independent parameters is $\sum_{i=1}^{p} m_i - p + pk - k(k - 1)/2$, which is the same as in the NOR approach. The parameters $\alpha_i^{(s)}$ and β_{ij} are unstandardized parameters. They can be standardized using Equations 23 and 24.

From Equation 28 it follows that:

$$\gamma_s^{(i)}(\mathbf{z}) = \frac{\exp[\alpha_s^{(i)} - \sum_{j=1}^{k} \beta_{ij} z_j]}{1 + \exp[\alpha_s^{(i)} - \sum_{j=1}^{k} \beta_{ij} z_j]} = \Psi[\alpha_s^{(i)} - \sum_{j=1}^{k} \beta_{ij} z_j] \quad , \tag{29}$$

where $\Psi(x)$ is the distribution function of the logistic distribution, see for example, Johnson and Kotz (1970; Chapter 22).

Equation 29 is very similar to Equation 21. The difference is between the normal $\Phi(x)$ and the logistic $\Psi(x)$ distribution functions. It is well known that these functions have very similar shapes. In fact, Lord and Novick (1968, p. 299) noted that

$$|\Phi(x) - \Psi(1.7x)| < 0.01 \quad \text{for all } x.$$

For the case of $k = 1$, Figure 3 shows four typical functions $\gamma_s^{(i)}(z)$ corresponding to the same parameter values used in Figure 1, the corresponding category response functions are shown in Figure 4. These functions have the same properties as those of NOR.

Let $\mathbf{x}_r = (x_1 = a_1, x_2 = a_2, \ldots, x_p = a_p)$ represent a full response pattern. Under the assumption of conditional independence, the conditional probability, for given $\mathbf{z}$, of the response pattern $\mathbf{x}_r$ is

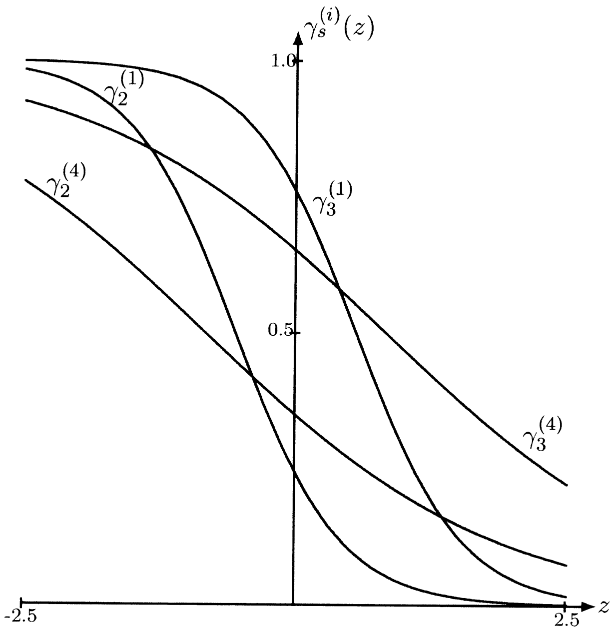

Figure 3
POM: Four Cumulative Response Functions $\gamma_s^{(i)}$

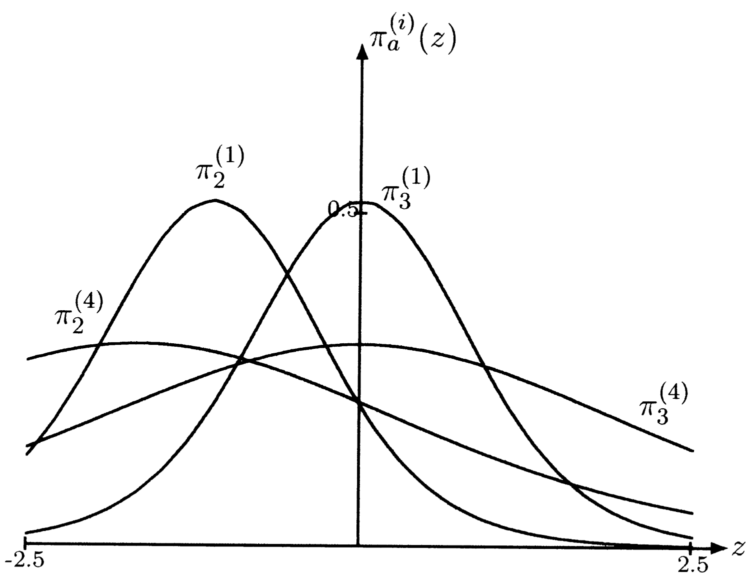

Figure 4
POM: Four Category Response Functions $\pi_a^{(i)}$

(30) $$\pi_r(\mathbf{z}) = \prod_{i=1}^{p} \pi_{a_i}^{(i)}(\mathbf{z}) = \prod_{i=1}^{p} [\gamma_{a_i}^{(i)}(\mathbf{z}) - \gamma_{a_i-1}^{(i)}(\mathbf{z})].$$

The unconditional probability π_r of the response pattern $\mathbf{x}_r$ is obtained by integrating $\pi_r(z)$ over the k-dimensional factor space:

(31) $$\pi_r(\theta) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \pi_r(\mathbf{z}) h(\mathbf{z}) d\mathbf{z}\ ,$$

where $h(\mathbf{z})$ is the density function of $\mathbf{z}$. As in the NOR approach, the latent factors are assumed to be independent with standard normal distributions, so that $h(\mathbf{z}) = \Pi_{j=1}^{k} \phi(z_j)$.

The integral in Equation 31 can be approximated to any practical degree of accuracy by Gauss-Hermite quadrature in the same way as for NOR, see Equation 27.

Full information maximum likelihood estimates may be obtained by minimizing the fit function (Equation 6) as for NOR. Alternatively, they can be obtained using an EM-algorithm described in Moustaki (2000). At the M-step of the algorithm one has to solve m_i non-linear equations in the one-factor model and m_i - 1 + k equations in the k-factor model. These non-linear equations are solved using a Newton-Raphson iterative scheme.

Although in principle one can have any number of factors, in practice it is difficult to use more than two factors because of the large number of quadrature points required in Equation 27. The program we use can fit up to two factors. The number of quadrature points is 48 for the one-factor model and 8 along each dimension, that is, 64 for the two-factor model.

Empirical Data Example

We will illustrate and compare the different approaches using both empirical and generated data. The analysis of an empirical data set is considered in this Section, whereas the analysis of generated data is considered in the section entitled “Generated Data Examples”.

The empirical data consists of the USA sample of the political action survey (Barnes & Kaase, 1979).[4] The variables are shown in Table 1. Permitted responses to these questions were *agree strongly* (***AS***), *agree* (**A**), *disagree* (**D**), *disagree strongly* (**DS**), *don’t know*, and *no answer*.

[4] The data was made available by the Zentralarchiv für Empirische Sosialforschung, University of Cologne. The data was originally collected by independent institutions in different countries. Neither the original collectors not the Zentralarchiv bear any responsibility for the analysis reported here.

Table 1
Empirical Data Variables

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 description

These six items have been used as operational indicators of the theoretical construct *political efficacy* in many surveys. Political efficacy has been defined as "*the feeling that individual political action does have, or can have, an impact upon the political process... The feeling that political and social change is possible, and that the individual citizen can play a part in bringing about this change*" (Campbell et al., 1954, p.187). The question is whether the six items measure one unidimensional factor.

The data we use is the listwise USA sample obtained after all cases with *don't know* or *no answer* responses have been eliminated. The listwise sample size is 1554.

Data Description

Each variable has four categories. Thus, there are 4096 possible response patterns but since we have data on only 1554 cases, every response pattern is not present in the data. Further screening of the data reveals that there are only 476 different response patterns present in the data. The 476 response patterns are distributed as shown in Table 2.

Table 2 shows that there are 284 different response patterns that occur only once, there are 83 different response patterns that occur twice, etc. There are 55 different response patterns with frequencies 6 or more. Further screening shows that among these there are 29 different response patterns occurring 10 or more times. These are distributed as shown in Table 3.

Table 2
Empirical Data: Number of Distinct Response Pattern with Small Frequencies

Frequency	No of Response Patterns	No of Cases
1	284	284
2	83	166
3	21	63
4	22	88
5	11	55
>5	55	898
Total	476	1554

It is seen in Table 3 that the most common response pattern is to answer *agree* to all six items (97 cases). The second most common response pattern is to answer *agree* to **COMPLEX** and *disagree* to all the other five items (70 cases). Note also that among those response patterns that occur 10 or more times there are no response *Disagree Strongly* on any of the six variables, that is, there are no 4's present in Table 3.

Parameter Estimates

We shall apply the UBN, NOR, and POM approaches to the full data set. There is a choice of reporting unstandardized or standardized parameters. Since most factor analysts are interested in standardized factor loadings, we report only standardized parameters. The UBN approach estimates standardized parameters directly. The NOR and POM approaches estimate unstandardized parameters directly. These have been transformed to standardized estimates using the formulas of Equations 23 and 24.

Although all three methods estimate both thresholds and factor loadings, thresholds are seldom of any interest. We therefore report only factor loadings here.

One Factor

The standardized factor loadings for the one-factor model are given in Table 4.

The NOR and POM loadings are generally larger than those of UBN. Except for **VOTING** and **COMPLEX**, the loadings of NOR and POM are

Table 3
Empirical Data: Distribution of Response Patterns with Large Frequencies.

Response Pattern	No of Cases
2 2 2 2 2 2	97
3 3 2 3 3 3	70
3 2 2 2 2 2	49
3 3 3 3 3 3	45
3 3 2 2 2 2	45
3 2 2 3 3 3	40
3 3 2 3 2 2	32
3 3 2 3 2 3	31
2 2 1 2 2 2	25
1 1 1 1 1 1	23
3 2 2 3 2 2	20
2 2 1 1 1 1	18
3 3 2 2 2 3	18
2 3 2 2 2 2	17
3 2 2 3 2 3	16
3 3 3 3 2 3	16
3 3 2 3 3 2	15
3 2 3 3 3 3	15
2 1 2 2 2 2	14
2 2 2 3 2 2	13
3 3 2 2 3 3	12
3 3 3 2 2 2	12
3 2 3 3 2 2	11
2 2 3 2 2 2	11
3 3 2 2 3 2	10
3 3 3 3 2 2	10
3 1 2 3 3 3	10
3 1 2 2 2 2	10
3 2 3 3 2 3	10

very close. For **VOTING** and **COMPLEX** the loadings are larger for POM than for NOR.

Since we do not know the “true” values we cannot say which method gives the “best” estimates. This issue will be considered in the section entitled “Generated Data Examples”.

Table 4
Standardized Loadings: One Factor

Item	UBN	NOR	POM
NOSAY	.60	.85	.82
VOTING	.37	.38	.58
COMPLEX	.52	.57	.75
NOCARE	.85	.93	.95
TOUCH	.77	.90	.92
INTEREST	.81	.91	.93

Two Factors

Since the two-factor solution is not uniquely defined, we must choose an identification condition to make the loadings comparable across methods. We choose to fix the second loading of **NOSAY** to zero. The solution determined in this way can be rotated to any orthogonal or oblique solution to facilitate easier interpretation, see, for example, Jöreskog (1969).

The standardized two-factor solutions are given in Table 5. Here the positive loadings of POM are larger than those of UBN and NOR, especially on the first factor. Otherwise, the three methods show similar patterns, but note the exceptionally large POM loadings of **VOTING** and **COMPLEX** on the first factor.

Table 5
Standardized Loadings: Two Factors

Item	UBN		NOR		POM	
NOSAY	.72	.00	.75	.00	.88	.00
VOTING	.46	-.07	.44	-.03	.70	-.13
COMPLEX	.52	.12	.65	.05	.75	.14
NOCARE	.75	.38	.81	.38	.85	.43
TOUCH	.57	.58	.54	.62	.68	.63
INTEREST	.62	.58	.60	.65	.72	.61

Testing the Model

Does the model fit the data? Does the two-factor model fit better than the one-factor model? Does the NOR model fit better than the POM model? In this section we consider ways of answering these questions.

Traditional factor analysts are used to test the model by comparing a fitted correlation or covariance matrix with a correlation or covariance matrix estimated from the sample without the model. To test the full model for ordinal variables one needs a test that compares the fitted probabilities of each response pattern with the corresponding sample proportions. Such a test can be based on the LR test in Equation 11 or the GF test in Equation 12 based on all the response patterns and their frequencies of occurrence and their expected frequencies under the model. These chi-square statistics can be used with NOR and POM since both of these maximize a likelihood function.

Table 6 gives the LR statistics for NOR and POM for one and two factors. The corresponding GF statistics for NOR and POM are given in Table 7. These are χ^2s with 444 degrees of freedom for the one-factor model and 439 degrees of freedom for the two-factor model.

These tables reveal two things: (a) When the model does not fit, the GF statistics can be much larger than the LR statistics. The reason for this will be clear later. However, when the model fits well, the LR and GF statistics are usually very close, and (b) The POM model fits better than the NOR model. Whether this is always the case will be investigated in the section entitled "Generated Data Examples".

Table 6
LR Statistics

	NOR	POM
One Factor	2192.13	1969.82
Two Factors	2040.36	1892.23

Table 7
GF Statistics

	NOR	POM
One Factor	30478962.77	897416.17
Two Factors	10262834.51	538221.25

The LR chi-square for NOR and POM are much too large for an acceptable fit. To get some further insight into the LR and GF statistics, one can compute the contribution from each response pattern. This will tell which response pattern is contributing maximally to the fit statistic. For the 29 most common response patterns of Table 3, that is, those that occur ten or more times, and for the two-factor solution of POM, these contributions are given in Table 8. In the computer we have the contributions of all the other 447 response patterns as well.

Note that the contributions to the LR statistics can be positive or negative. A positive contribution means that the model underestimates the observed proportions. This is the case for most of the common response patterns. A negative contribution means that the model overestimates the observed proportion. An advantage with the use of the LR contributions is that one can see directly whether over or underestimation occurs. For the GF contributions this is not possible; one would have to compare the expected and observed frequencies to tell whether this occurs.

Among the 29 most common response patterns, the largest positive LR contribution is for the pattern (2 2 2 2 2 2) and the largest negative contribution is for the pattern (2 3 2 2 2 2). It turns out that these are the two largest LR contributions among all the 476 response patterns in the sample. Further screening of the file containing all the 476 LR and GF contributions, reveals that the largest LR contributions occur for the most frequent response patterns, that is, those in Table 8.

The GF chi-squares in Table 7 are extremely large. The reason for this can be seen as follows. The response patterns in Table 8 were sorted in descending order according to the observed frequencies. Suppose instead we sort the 476 response patterns in ascending order according to the expected frequencies. The ten smallest expected frequencies, the corresponding observed frequencies and their chi-square contributions are shown in Table 9.

Among the 476 response patterns, the largest contribution to GF Chi-square is 368129.79 which is obtained for the pattern (2 4 4 4 1 1) which occur only once in the data. The expected frequency of this pattern under the model is 0.00000272. Further screening of the file reveals that all the large contributions to the GF statistic occur for patterns with very small probabilities. This suggest that the GF chi-square derives most of its value from patterns with small probabilities, whereas the LR chi-square derives most of its from patterns with large probabilities. This is a bad property of the GF statistic.

The LR and GF chi-squares in Tables 6 and 7 are grossly distorted due to sparseness of the contingency table. The data can be viewed as

Table 8
Contributions to Fit Statistics: Large LR-Contributions

r	n_r	$N\hat{\pi}_r$	$2Np_r \ln(p_r / \hat{\pi}_r)$	$N(p_r - \hat{\pi}_r)^2 / \hat{\pi}_r$	Pattern
1	97	36.15	191.48	102.42	2 2 2 2 2 2
2	70	36.34	91.79	31.19	3 3 2 3 3 3
3	49	37.91	25.14	3.24	3 2 2 2 2 2
4	45	25.50	51.12	14.91	3 3 3 3 3 3
5	45	34.23	24.62	3.39	3 3 2 2 2 2
6	40	28.30	27.67	4.83	3 2 2 3 3 3
7	32	25.96	13.40	1.41	3 3 2 3 2 2
8	31	24.17	15.43	1.93	3 3 2 3 2 3
9	25	14.85	26.05	6.94	2 2 1 2 2 2
10	23	12.21	29.12	9.53	1 1 1 1 1 1
11	20	20.88	-1.72	0.04	3 2 2 3 2 2
12	18	9.43	23.27	7.79	2 2 1 1 1 1
13	18	11.14	17.26	4.22	3 3 2 2 2 3
14	17	21.08	-7.32	0.79	2 3 2 2 2 2
15	16	18.98	-5.47	0.47	3 2 2 3 2 3
16	16	13.15	6.27	0.62	3 3 3 3 2 3
17	15	11.21	8.73	1.28	3 3 2 3 3 2
18	15	15.64	-1.25	0.03	3 2 3 3 3 3
19	14	16.34	-4.33	0.34	2 1 2 2 2 2
20	13	9.89	7.10	0.98	2 2 2 3 2 2
21	12	6.35	15.27	5.02	3 3 2 2 3 3
22	12	9.89	4.64	0.45	3 3 3 2 2 2
23	11	7.28	9.08	1.90	3 2 3 3 2 2
24	11	5.77	14.19	4.74	2 2 3 2 2 2
25	10	6.08	9.94	2.52	3 3 2 2 3 2
26	10	10.77	-1.47	0.05	3 3 3 3 2 2
27	10	7.22	6.52	1.07	3 1 2 3 3 3
28	10	13.09	-5.38	0.73	3 1 2 2 2 2
29	10	8.46	3.34	0.28	3 2 3 3 2 3

a 4 × 4 × 4 × 4 × 4 × 4 contingency table with 4096 cells but the sample size is only 1554. Sparseness of a contingency table is often discussed in terms of small expected cell frequencies due to the fact that the sample size is smaller than the number of cells. The distorting effect of sparseness on the chi-square test is well known, see for example, Agresti and Yang (1987) and

Table 9
Contributions to Fit Statistics: Large GF-Contributions

r	n_r	$N\hat{\pi}_r$	$2Np_r \ln(p_r / \hat{\pi}_r)$	$N(p_r - \hat{\pi}_r)^2 / \hat{\pi}_r$	Pattern
1	1	0.00000272	25.63	368129.79	2 4 4 4 1 1
2	1	0.00001415	22.33	70651.28	1 2 4 4 1 3
3	1	0.00001986	21.65	50355.55	4 4 1 3 1 4
4	1	0.00007528	18.99	13281.37	1 1 4 4 4 4
5	1	0.00015620	17.53	6400.07	4 4 4 1 1 1
6	1	0.00023514	16.71	4250.74	4 4 4 4 3 1
7	1	0.00027041	16.43	3696.10	1 1 3 4 4 4
8	1	0.00044807	15.42	2229.82	4 4 3 4 2 1
9	1	0.00047435	15.31	2106.13	1 4 4 1 1 2
10	1	0.00054585	15.03	1830.01	3 4 1 4 2 4

Agresti (1990, pp 246-250). Reiser and VandenBerg (1994) investigated the effect of sparseness of the LR and GF statistics for the one-factor model with dichotomous variables.

We consider two alternative ways of reducing the distorting effects on chi-square.

Alternative 1

Suppose we calculate the sum of the LR and GF contributions only over those response patterns whose expected frequency exceeds a value ν and use these as test statistics. The LR and GF chi-squares in Tables 6 and 7 are based on $\nu = 0$ but values of 1, 2, 5, and 10 of ν have been suggested in the literature. For the two-factor POM model the results are shown in Table 10. The χ^2s in this table have been computed as follows. The response patterns whose expected frequencies exceeds ν are selected. The number of such response patterns is given in the second row of the table and the sum of their observed frequencies are given in the third row of the table. The degrees of freedom is the number of response patterns minus 1 minus the number of independent parameters (29 in this case). Note that for $\nu = 10$ the degrees of freedom is negative. So this value is too large. The expected frequencies are multiplied by a constant such that their sum equals the sum of observed frequencies (i.e. the number of observations). The LR- and GF-statistics

Table 10
Fit Measures based on Subsamples

	ν = 1	ν = 2	ν = 5	ν = 10
Number of response patterns	232	162	65	22
Number of observations	1274	1164	916	624
Degrees of Freedom	203	133	36	-7
LR-χ^2	316.02	292.65	179.65	84.30
GF-χ^2	321.24	295.46	181.75	89.74

are then computed by the formulas of Equations 11 and 12. It should be pointed out that these are not real chi-squares since the model has not been fitted on the subset of observations. If the model is fitted on the subset of observations, the chi-squares will be smaller than those reported in Table 10.

The "chi-squares" of Table 10 are much much smaller than those in Tables 6 and 7 and the LR- and GF-statistics are quite close. Hence, it is clear that the sparseness of the contingency table has an estremely adverse effect on the chi-square statistics in this data set. However, even the values in Table 10 suggest that the two-factor POM model does not fit the data.

Alternative 2

Suppose we regard the 476 response patterns as categories of a multinomial distribution. The problem is that many of these categories have very small expected frequencies. In fact, 244 of the 476 categories have expected frequencies less than 1. The usual way of dealing with this problem is to combine categories in such a way that all retained categories have expected frequencies exceeding a value ν, for example, 5, 10 or 20. The LR- and GF-statistics can then be calculated on the bases of this reduced set of categories. For the two-factor POM solution and for values of ν equal to 5, 10, 15, 20, 25, and 30, these LR- and GF-statistics are given in Table 11.

The second column *cat* is the number of categories in the reduced set. The third column *df* is the degrees of freedom. The chi-squares of Table 11 are also much much smaller than those in Tables 6 and 7 and the LR-statistic is much smaller than the GF-statistics. However, even the values in Table 11 suggest that the two-factor POM model does not fit the data.

While this kind of analysis will reduce the distorting effects on the chi-square statistics and reveal the resonse patterns that cause the most lack of fit, it is not

Table 11
Fit Measures based on Combined Categories

ν	*cat*	*df*	χ^2_{LR}	χ^2_{GF}
5	159	130	924.83	3160.65
10	103	74	862.69	2887.90
15	76	47	884.26	5453.48
20	62	33	759.82	2052.87
25	52	23	718.99	1807.95
30	45	16	690.53	1606.85

useful in deciding what should be done to improve the fit of the model. From a practical point of view there are two things that can be done.

1. Reduce the number of categories. In this case one could collapse the two categories *agree strongly* and *agree* into one category and the two categories *disagree strongly* and *disagree* into one category. This can be done for any of the variables. If it is done for all variables, all variables will be dichotomous. The total number of possible response patterns will then be 64, but actually only 62 occur in the sample.

2. Eliminate the most offending variables thereby obtaining more homogeneity for the retained variables. In this case we would recommend deleting the variable **VOTING** since this has an ambiguous question wording and is the variable with the smallest factor loadings.

Since this is a methodological article rather than substantive, we shall not pursue these matters here but simply state that both of these ways will improve the fit considerably.

Measurement of Fit

If we focus on measurement of fit rather than the testing of the model, another way of measuring fit emerges. This measures fit relative to the sample univariate and bivariate marginal distributions. For the efficacy data the univariate marginal distributions are given in Table 12.

The bivariate marginal distributions are represented by 15 contingency tables of order 4 × 4, one for each pair of variables. These contingency tables are given in Table 13.

For each cell of these univariate and bivariate contingency tables, we define a LR-Fit and a GF-Fit. We use these terms to emphasize that there are no χ^2-distribution associated with them.

Table 12
Univariate Marginal Distributions

	Frequency				Percentage			
	AS	**A**	**D**	**DS**	**AS**	**A**	**D**	**DS**
NOSAY	160	471	804	119	10.3	30.3	51.7	7.7
VOTING	271	635	576	72	17.4	40.9	37.1	4.6
COMPLEX	309	881	304	60	19.9	56.7	19.6	3.9
NOCARE	238	635	626	55	15.3	40.9	40.3	3.5
TOUCH	257	832	440	25	16.5	53.5	28.3	1.6
INTEREST	255	719	551	29	16.4	46.3	35.5	1.9

For category a of variable g the LR- and GF-Fits are defined as follows

(32) $$\mathrm{LR-Fit}_a^{(g)} = 2Np_a^{(g)} \ln[p_a^{(g)} / \hat{\pi}_a^{(g)}],$$

(33) $$\mathrm{GF-Fit}_a^{(g)} = N[p_a^{(g)} - \hat{\pi}_a^{(g)}]^2 / \hat{\pi}_a^{(g)},$$

Summing these over a gives the univariate LR- and GF-Fits for variable g. Similarly we define bivariate LR- and GF-Fits for category a of variable g and category b of variable h as

(34) $$\mathrm{LR-Fit}_{ab}^{(gh)} = 2Np_{ab}^{(gh)} \ln[p_{abb}^{(gh)} / \hat{\pi}_a^{(gh)}],$$

(35) $$\mathrm{GF-Fit}_{ab}^{(gh)} = N[p_{ab}^{(gh)} - \hat{\pi}_{ab}^{(gh)}]^2 / \hat{\pi}_{ab}^{(gh)},$$

Summing these over a and b gives the bivariate LR- and GF-Fits for variable g and h. The probabilities $\hat{\pi}$ can be evaluated under each model UBN, NOR and POM. Since these fit measures are based on different number of cells for different variables and different pairs of variables, we divide them by the number of cells to make them comparable across variables and pairs of variables.

Reiser (1996) considered similar ideas for the case of dichotomous variables and suggested using residuals for each cell of the univariate and bivariate marginals. These residuals are simply the square roots of our GF-Fit in Equations 33 and 35, see Cochran (1954). The asymptotic variance of these residuals are given in Rao (1965, eq. 6b.3.3). This involves the information matrix and is therefore complicated to compute. Reiser (1996) suggested

Table 13
Bivariate Marginal Distributions

	VOTING				COMPLEX				NOCARE			
NOSAY	**AS**	**A**	**D**	**DS**	**AS**	**A**	**D**	**DS**	**AS**	**A**	**D**	**DS**
AS	69	51	27	13	83	52	16	9	80	58	19	3
A	80	297	85	9	120	287	55	9	104	278	88	1
D	92	275	413	24	89	491	201	23	44	281	455	24
DS	30	12	51	26	17	51	32	19	10	18	64	27

	TOUCH				INTEREST			
NOSAY	**AS**	**A**	**D**	**DS**	**AS**	**A**	**D**	**DS**
AS	75	67	16	2	84	50	24	2
A	99	311	60	1	99	281	91	0
D	62	411	316	15	57	357	378	12
DS	21	43	48	7	15	31	58	15

	COMPLEX				NOCARE				TOUCH			
VOTING	**AS**	**A**	**D**	**DS**	**AS**	**A**	**D**	**DS**	**AS**	**A**	**D**	**DS**
AS	102	125	37	7	73	108	78	12	82	134	49	6
A	131	389	99	16	97	327	203	8	100	398	133	4
D	63	341	151	21	55	190	311	20	58	275	233	10
DS	13	26	17	16	13	10	34	15	17	25	25	5

	INTEREST			
VOTING	**AS**	**A**	**D**	**DS**
AS	84	104	74	9
A	105	351	177	2
D	50	246	269	11
DS	16	18	31	7

	NOCARE				TOUCH				INTEREST			
COMPLEX	**AS**	**A**	**D**	**DS**	**AS**	**A**	**D**	**DS**	**AS**	**A**	**D**	**DS**
AS	133	115	56	5	131	142	31	5	122	141	41	5
A	84	418	363	16	90	521	263	7	101	444	326	10
D	16	88	184	16	23	151	124	6	24	118	159	3
DS	5	14	23	18	13	18	22	7	8	16	25	11

	TOUCH				INTEREST			
NOCARE	**AS**	**A**	**D**	**DS**	**AS**	**A**	**D**	**DS**
AS	150	73	13	2	154	70	14	0
A	84	466	83	2	79	437	117	2
D	20	280	315	11	17	200	398	11
DS	3	13	29	10	5	12	22	16

	INTEREST			
TOUCH	**AS**	**A**	**D**	**DS**
AS	167	70	19	1
A	80	531	216	5
D	7	112	306	15
DS	1	6	10	8

using the residuals divided by the square root of their asymptotic variance as measures of cell fit and then summing these over all cells in the univariate and bivariate marginal to define a fit measure for each variable and each pair of variables. While such measures of fit might have a better statistical basis, the measures we propose are much easier to compute. As a simple guide line we suggest that values larger than 4 are indicative of poor fit.

The univariate and bivariate LR-Fits for the two-factor solutions of UBN, NOR, and POM are given in Table 14 and the corresponding GF-Fits are given in Table 15. The univariate fits are in the diagonal and the bivariate fits are below the diagonal. The overall fit measure is the average of all the off-diagonal fits.

From these tables the following observations can be made:

1. The GF-Fits are generally larger than the LR-Fits. The reason for this is that the GF-Fit gets too large contributions from cells with small fitted probabilities $\hat{\pi}$. For this reason we prefer to use the LR-Fits.

2. The univariate margins fit much better than the bivariate margins.

3. The fit seems to be better for POM than for NOR and UBN. But note that the overall LR fit for UBN will always be smaller than for that of NOR because this is minimized under UBN and UBN and NOR makes the same assumptions (underlying bivariate normality). The same is not true for UBN and POM since these use different assumptions (normal vs. logistic).

4. The fit is not adequate for any of the models. In particular, the contingency table for **NOSAY** and **VOTING** fits particularly bad.

One can investigate the reason for the poor fit for **NOSAY** and **VOTING** by examining the individual cell contributions to the LR-Fit. We do this for the two-factor UBN solution only. These are given in Table 16. This reveals that the fit can be improved by collapsing the **D** and **DS** categories. To see this just add the numbers in the rows and columns corresponding to **D** and **DS**.

The total LR-Fit, 216.05, is then divided by 16 since there are 16 cells in Table 16. This gives the value 13.5 given in Table 14.

We summarize the measurement of fit by giving the overall LR-Fits for each method and for one and two factors. These are given in Table 17.

It is clear that, no matter what measure of fit is used, LR or GF, the model does not fit the data. What can one do about this? Are there ways to improve the fit of the model?

In classical factor analysis one would just add factors until a satisfactory fit is obtained. With six observed variables, one would obtain a perfect fit for three factors since the classical degrees of freedom is zero for three factors. This approach does not work for ordinal variables. In this case, for the three-factor model, although very difficult to estimate with NOR and POM, all fit measures are only slightly smaller than those of the two-factor model. The three-factor model will also be rejected. The reason for this is that the lack of fit is not due to the factor model. Rather, it is due to the other assumptions in the model. These other assumptions are: (a) factors are independent and normally distributed, (b) given the factors the responses to different variables are independent (conditional independence), and (c) the cumulative response function is the normal distribution (NOR) or the logistic distribution (POM).

Table 14
Univariate and Bivariate LR-Fits

			UBN			
NOSAY	0.0					
VOTING	13.5	0.0				
COMPLEX	4.9	2.9	0.0			
NOCARE	5.0	5.0	4.0	0.0		
TOUCH	5.3	3.8	5.5	5.6	0.0	
INTEREST	5.0	4.8	4.0	7.3	5.6	0.0
OVERALL LR-FIT	**5.5**					

			NOR			
NOSAY	0.2					
VOTING	13.8	0.0				
COMPLEX	4.9	2.7	0.1			
NOCARE	5.2	5.3	4.1	0.9		
TOUCH	5.5	4.0	5.6	5.6	0.6	
INTEREST	5.1	4.9	4.1	7.3	5.7	0.5
OVERALL LR-FIT	**5.6**					

			POM			
NOSAY	0.1					
VOTING	11.5	0.0				
COMPLEX	4.0	2.4	0.1			
NOCARE	3.5	4.6	3.2	0.4		
TOUCH	4.5	3.2	4.8	3.8	0.1	
INTEREST	4.2	4.3	3.5	5.2	3.8	0.2
OVERALL LR-FIT	**4.4**					

Table 15
Univariate and Bivariate GF-Fits

	UBN					
NOSAY	0.0					
VOTING	19.8	0.0				
COMPLEX	8.9	3.1	0.0			
NOCARE	11.8	7.2	4.7	0.0		
TOUCH	8.5	4.7	8.4	36.1	0.0	
INTEREST	7.4	6.3	5.4	59.3	28.3	0.0
OVERALL GF-FIT	**14.7**					

	NOR					
NOSAY	0.2					
VOTING	16.9	0.0				
COMPLEX	8.0	2.9	0.0			
NOCARE	9.2	7.0	4.8	0.2		
TOUCH	8.6	4.6	7.5	26.6	0.1	
INTEREST	7.1	5.8	4.8	47.9	16.2	0.1
OVERALL GF-FIT	**11.9**					

	POM					
NOSAY	0.1					
VOTING	15.9	0.0				
COMPLEX	5.9	2.6	0.1			
NOCARE	5.9	6.0	3.3	0.4		
TOUCH	7.5	3.9	6.3	8.5	0.1	
INTEREST	5.9	5.4	4.0	15.4	5.2	0.2
OVERALL GF-FIT	**6.8**					

Assumption 1 cannot be tested separately. There is no way we can have any information about the distribution of the latent factors. Any assumption about their distribution is as good as any other. Assumption 2 should not be questioned since this is the central part of the definition of factor analysis. The latent factors are supposed to account for the dependencies among the

Table 16
Contribution to LR-Fits: **NOSAY** BY **VOTING**

	AS	**A**	**D**	**DS**
AS	27.90	-32.98	-7.57	58.02
A	-47.51	197.03	-87.23	-2.59
D	-16.91	-78.19	164.42	-29.33
DS	95.39	-25.04	-24.21	24.86

Total LR-Fit = **216.05**

Table 17
Total LR-Fits

	UBN	NOR	POM
One Factor	5.7	7.4	4.5
Two Factors	5.5	5.6	4.4

observed variables. The only assumption that can be modified is Assumption 3. The analysis of the Efficacy data suggests that neither the normal nor the logistic distribution holds. These distributional assumptions are as arbitrary as any other similar assumption but they are the only two that have been considered in the literature.

Generated Data Examples

The disadvantage of using empirical data to compare different approaches is that the true parameter values are not known so that one does not know which set of estimates are closest to the true values. By using generated data one can study the parameter estimates in terms of bias and mean squared error. Unfortunately, because of the enormously long time it takes to estimate factor models with NOR and POM, we are unable to perform a full scale simulation using many replicates. We shall therefore analyze only two samples under each condition. Nevertheless, we think these examples give some information regarding the behavior of the LR-χ^2 and the LR-Fit measures when the model holds.

Data for the NOR and POM models can be generated as follows:

1. Specify the sample size N and standardized parameters $\tau_a^{(i)}$ and λ_{ij}. Transform these to unstandardized parameters $\alpha_a^{(i)}$ and β_{ij}.

2. For each of the $Q = \prod_{i=1}^{p} m_i$ possible response patterns r compute the corresponding probability π_r at the true unstandardized parameters under NOR or POM. Then compute the cumulative probabilities $\nu_s = \sum_{r=1}^{s} \pi_r$, $s = 1, 2, \ldots, Q$.

3. Draw a uniform random number u and generate response pattern r if $\nu_{r-1} < u \leq \nu_r$.

4. Repeat Step 3 N times.

In this section we analyze data generated from two different populations, one small and one large. Since POM will be favored if the data is generated under POM and NOR will be favored if the data is generated under NOR, we generate two samples, one under NOR and one under POM and analyze each by UBN, NOR, and POM. The two samples are generated using the same random numbers u. What makes the two samples different is that each of the Q possible response patterns have a different probability under NOR and POM.

Small Population

The small population consists of four variables satisfying a one-factor model. Each variable has five categories. Thus, there are Q = 625 possible response pattern. The population parameters for the small population is given in Table 18. The model has 20 parameters. The sample size is N=800. In the NOR-generated data there are 163 distinct response patterns in the sample. In the POM-generated data there are 305 distinct response patterns in the sample. Here we see one difference between NOR and POM. POM gives more probability to a wider set of response patterns, whereas NOR concentrates more probabilities to a smaller set of response patterns.

Table 18
Small Population: True Standardized Parameters

Variable	Thresholds				Loadings
i	τ_{i1}	τ_{i2}	τ_{i3}	τ_{i4}	λ_i
1	-1.50	-0.50	0.50	1.50	0.90
2	-2.50	-0.50	0.50	1.00	0.80
3	-1.00	-0.50	0.50	2.50	0.70
4	-1.50	-0.50	0.50	1.50	0.60

The estimated factor loadings are shown in Table 19 which shows the following.

1. For NOR-generated data, the UBN and NOR estimates are closer to the true values than the POM estimates. POM overestimates the loadings.

2. For POM-generated data, the POM estimates are closer to the true values than the UBN and NOR estimates. The latter underestimates the true values.

Results concerning LR-χ^2s are shown in Table 20.

Line 3 of Table 20 gives the LR-χ^2s based on all the response patterns in the sample (163 for the NOR-generated data and 305 for the POM-generated data). The degrees of freedom is given after the slash. If one were to use these LR-χ^2 to test the model, the model would be rejected even though it holds. Thus there is some distortion of chi-square due to sparseness also for this data even though the sample size (800) exceeds the total number of cells (625). However, with generated data it is sufficient to use a very small value of ν to eliminate the distortion. Line 4 of Table 20 gives the LR-χ^2s for all response patterns whose expected frequencies exceed $\nu = 0.05$

Table 19
Small Population: Estimated Factor Loadings

	NOR-Generated Data			POM-Generated Data		
True	UBN	NOR	POM	UBN	NOR	POM
.90	.90	.85	.90	.81	.72	.88
.80	.82	.82	.88	.61	.65	.82
.70	.71	.72	.89	.46	.48	.69
.60	.57	.58	.77	.31	.32	.51

Table 20
Small Population: LR-χ^2

NOR-Generated Data		POM-Generated Data	
NOR	POM	NOR	POM
293.10/142	374.47/142	492.05/284	491.85/284
156.45/141	165.44/141	222.61/283	222.85/283

computed according to Alternative 1. These chi-squares are all non-significant in agreement with the fact that the model holds. There was only one response pattern that caused the distortion.

The LR-Fits are shown in Table 21. Here it is seen that UBN and NOR fits much better than POM for NOR-generated data but for POM-generated data there is not much difference. These numbers are all well below the critical value 4, thus indicating a good fit of the model.

Large Population

The large population consists of eight variables satisfying a two-factor model. Each variable has four categories. Thus, there are $Q = 4^8 = 65536$ possible response patterns. The population parameters for the large population are given in Table 22. The model has 39 parameters, 24 thresholds

Table 21
Small Population: LR-Fit

NOR-Generated Data			POM-Generated Data		
UBN	NOR	POM	UBN	NOR	POM
0.45	0.53	1.56	0.70	0.73	0.74

Table 22
Large Population: True Standardized Parameters

Variable	Thresholds			Loadings	
i	τ_{i1}	τ_{i2}	τ_{i3}	λ_{i1}	λ_{i2}
1	-1.20	0.00	1.20	0.8	0.0
2	-1.20	0.00	1.20	0.7	0.1
3	-1.20	0.00	1.20	0.6	0.2
4	-1.20	0.00	1.20	0.5	0.3
5	-1.20	0.00	1.20	0.4	0.4
6	-1.20	0.00	1.20	0.3	0.5
7	-1.20	0.00	1.20	0.2	0.6
8	-1.20	0.00	1.20	0.1	0.7

and 15 factor loadings. The sample size is $N = 1000$. In the NOR-generated data there are 893 distinct response patterns. In the POM-generated data there are 984 distinct response patterns. Thus, in the POM-generated data, nearly all of the 1000 possible response patterns are present in the sample. This is in agreement with the statement we made in connection with the small data example, namely that POM gives more probability to a wider set of response patterns, whereas NOR concentrates more probabilities to a smaller set of response patterns.

The estimated factor loadings are given in Table 23 for the first factor and in Table 24 for the second factor. We look first at the first factor. For the NOR generated data the UBN estimates seems to be closest to the true values. NOR and POM seems to overestimate the loadings. For the POM generated data, the POM estimates seems to be closest to the true values. For the second factor the situation seems to be exactly the same as for the first factor.

The LR χ^2-statistics are given in Table 25 for $\nu = 0$ and $\nu = 0.05$. Those for $\nu = 0$ are quite distorted but it is seen that also in this case a small value of ν is sufficient to eliminate the distortions.

The LR-Fits are given in Table 26. They all indicate that the model fits the uni- and bivariate margins quite well. In terms of LR-Fits, the POM model fits better than the NOR model regardless of how the data is generated but for the POM-generated data the differences in fit are very small. As previously stated UBN will always fit better than NOR for NOR-generated data.

Table 23
Large Population: Estimated Factor Loadings for Factor One

	NOR-Generated Data			POM-Generated Data		
True	UBN	NOR	POM	UBN	NOR	POM
.80	.83	.91	.94	.66	.67	.82
.70	.66	.63	.84	.49	.49	.70
.60	.57	.54	.74	.39	.39	.58
.50	.49	.48	.67	.39	.45	.60
.40	.40	.39	.53	.20	.21	.32
.30	.29	.29	.39	.19	.20	.29
.20	.22	.23	.31	.08	.09	.13
.10	.07	.09	.10	.03	.04	.06

Table 24
Large Population: Estimated Factor Loadings for Factor Two

	NOR-Generated Data			POM-Generated Data		
True	UBN	NOR	POM	UBN	NOR	POM
.00	.00	.00	.00	.00	.00	.00
.10	.09	.12	.10	.06	.05	.06
.20	.25	.30	.32	.20	.19	.27
.30	.26	.33	.36	.11	.09	.14
.40	.43	.50	.58	.27	.26	.42
.50	.52	.61	.70	.36	.41	.53
.60	.56	.54	.73	.41	.39	.61
.70	.74	.74	.89	.42	.42	.61

Table 25
Large Population: LR-χ^2

NOR-Generated Data		POM-Generated Data	
NOR	POM	NOR	POM
4845.75/853	4798.00/853	7709.47/944	7700.61/944
502.49/541	493.23/570	12.92/108	14.29/106

Table 26
Large Population: LR-Fit

NOR-Generated Data			POM-Generated Data		
UBN	NOR	POM	UBN	NOR	POM
0.51	0.96	0.51	0.48	0.50	0.47

These analysis of two sets of generated data suggests that the methods work for estimating factor loadings but the LR-χ^2 cannot be trusted to be used to test the model unless one eliminates the response patterns with very small expected frequencies.

Summary and Discussion

We set out to compare four approaches, UMN, UBN, NOR, and POM. *A priori*, the UMN, NOR, and POM approaches should give better estimates and better fit than the UBN approach because they are full information methods that use all information in the data, whereas UBN is a limited information method that uses only information from the univariate and bivariate marginal distributions. UMN, NOR, and POM are real maximum likelihood methods in that they each maximize a single likelihood function. The UBN approach has a weaker theoretical foundation as it maximizes the sum of all univariate and bivariate likelihoods and these likelihoods are not independent. The advantage of UBN compared to existing approaches (as available in programs for structural equation modeling) is that it avoids the stepwise nature of these approaches and the use of polychoric correlations and their asymptotic covariance matrix.

We found that the UMN approach is computationally unfeasible for more than four variables. Since the NOR and POM approaches are computationally much heavier than UBN, the question arises: is it worth the trouble to use these methods or does one obtain almost as good results with UBN?

Parameter Estimates

For the empirical data set the factor loading estimates were rather similar across approaches, but there was a tendency for POM to estimate larger loadings for both one and two factors. The two sets of generated data suggest that factor loadings are fairly close across different approaches. In fact, differences between approaches do not seem to be of practical importance.

Goodness of Fit

In theory correct goodness of fit tests are available for NOR and POM using the LR and GF χ^2-statistics. However, as demonstrated both by the empirical data set and by the generated data sets, there are serious problems with these LR and GF χ^2-statistics if these are based on all response patterns ocurring in the sample due to too small expected frequencies for some response patterns. This problem is likely to be serious for data where the sample size is small relative to the number of response patterns, a common situation in practice. We found that this problem is much more serious for the GF statistic than for the LR statistic. The problem can be reduced to some extent by computing the χ^2 -statistic only for response patterns with expected frequencies exceeding a certain value ν, but the arbitrariness of the choice of ν is unsatisfactory and the asymptotic chi-square distribution is no longer guaranteed. Another possibility to solving this problem would be to estimate the model only over those response patterns that occur five or more times, say. This might give a correct chi-square value but at the expense of throwing away some data.

To measure overall fit to the data we proposed measuring fit relative to the univariate and bivariate margins. These fit measures can be used with UBN, NOR, and POM. For these LR-Fits, there are very small differences between the three approaches. For the empirical data set the examination of detailed fit for each pair of variables, revealed the same pair of variables as the major cause of lack of fit. For deciding how to improve the fit of the model these LR-Fits are more useful than the LR-χ^2. But, unfortunately these LR-Fits do not have a known distributions under the model so one has to rely on arbitrary guide lines. Further studies are needed to investigate this.

POM seems to fit the data better than NOR. Perhaps this is because POM gives more probability to a wider range of response patterns. We were only able to demonstrate one case where POM fitted significantly worse than NOR.

Feasibility

The NOR and POM approaches are computationally heavy. In each iteration one has to go through every distinct response pattern in the data and estimate its likelihood according to the model. This likelihood is not available in closed form but can only be approximated by numerical integration requiring several quadrature points. The NOR and POM methods work reasonably well for one and two factors. If an accurate solution is not required, it is possible to handle three and four factors by reducing the number of quadrature points along each dimension. But as the number of factors increases the methods become too heavy even on today's fast computers. POM is slightly faster than NOR. If one has to choose between POM and NOR, one should choose POM.

By contrast, the UBN method is quite feasible with many variables and factors. It is no problem to handle 40 variables and 10 factors, albeit one may need a large sample to obtain good estimates. The generated data examples demonstrated that the UBN method usually gives good estimates.

Table 27 illustrates how computer time depends on number of variables, number of categories, number of factors, and sample size. The notation used is 0 = time is constant, + = time increases approximately linearly, ++ = time increases slightly more than linearly, and +++ = time increases approximately exponentially.

One particular advantage with UBN, relative to POM and NOR, is that computer time does not increase with sample size.

Conclusion

In theory NOR and POM are better methods than UBN but from a practical point of view their uses are very limited. NOR and POM suffers from the fact that it is difficult to carry out a full scale simulation study to evaluate their properties. Further research is needed in numerical analysis to make the required integrals easier and faster to compute. It is also desireable to define other and more flexible cumulative response functions than the normal and the logistic.

Most factor analysts want to analyze many variables with many factors. Such analysts have no choice but to use the UBN method or some similar method such as weighted least squares based on polychoric correlations available in PRELIS/LISREL and MPLUS. The UBN method might not fit the data because underlying bivariate normality does not hold. This will also be the case with PRELIS/LISREL and MPLUS. Underlying bivariate normality can be tested separately with PRELIS. Further research is needed on the effects of underlying non-normality on parameter estimates and fit measures.

Table 27
Time to Compute the Estimates

	UBN	NOR	POM
Variables	+	++	++
Categories	+	+	+
Factors	+	+++	+++
Sample Size	0	++	++

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Agresti, A. & Yang, M. C. (1987). An empirical investigation of some effects of sparseness in contingency tables. *Computational Statistics and Data Analysis*, *17*, 9-21.

Arminger, G. & Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 51-73). New York: Plenum Press.

Barnes, S. H. & Kaase, M. (Eds.) (1979). *Political action: Mass participation in five western democracies.* Beverley Hills and London: Sage Publications.

Bartholomew, D. J. & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London: Arnold Kendall's Library of Statistics 7.

Campbell, A., Gurin, G., & Miller, W. A. (1954). *The voter decides.* New York: Wiley.

Cochran, W. G. (1954). Some methods for strengthening the common chi-square tests. *Biometrical Journal*, *10*, 417-451.

Genz, A. & Keister, B. (1996). Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight. *Journal of Computational and Applied Mathematics*, *71*, 299-309.

Johnson, N. L. & Kotz, S. (1970). *Distributions in statistics: Continuous univariate distributions1*. New York: John Wiley & Sons.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.

Jöreskog, K. G. (1979). Basic ideas of factor and component analysis. In K. G. Jöreskog & D. Sörbom (Eds.), *Advances in factor analysis and structural equation models* (pp. 5-20). Cambridge, MA: Abt Books.

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

Lee, S. Y., Poon, W. Y., & Bentler, P. (1990). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics and Probability Letters*, *9*, 91-97.

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

Lord, F. M. & Novick, M. E. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley Publishing Co.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, *42*, 109-142.

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.

Moustaki, I. (2000). A latent variable model for ordinal variables. *Applied Psychological Measurement*, *24*, 211-224.

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

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variables indicators. *Psychometrika*, *49*, 115-132.

Muthen, B. & Satorra, A. (1995). Technical aspects of Muthen's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, *60*, 489-503.

Reiser, M. (1996). Analysis of residuals for the multinomial item response model. *Psychometrika*, *61*, 509-528.

Reiser, M. & VandenBerg, M. (1994). Validity of the chi-square test in dichotomous variable factor analysis when expected frequencies are small. *British Journal of Mathematical and Statistical Psychology*, *47*, 85-107.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement No. 17*.

Schervish, M. J. (1984). Algorithm AS 195: Multivariate normal probabilities with error bound. *Applied Statistics*, *3*, 81-94.

Takane, Y. & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393-408.

Thissen, D. (1991). *MULTILOG: Multiple, categorical items analysis and test scoring using item response theory* [Computer program]. Chicago: Scientific Software, Inc.

Accepted July, 2000.