PSYCHOMETRIKA—VOL. 66, NO. 4, 515–530
DECEMBER 2001

FACTOR ANALYSIS WITH (MIXED) OBSERVED AND LATENT VARIABLES IN THE EXPONENTIAL FAMILY

MICHEL WEDEL

UNIVERSITIES OF GRONINGEN AND MICHIGAN

WAGNER A. KAMAKURA

DUKE UNIVERSITY

We develop a general approach to factor analysis that involves observed and latent variables that are assumed to be distributed in the exponential family. This gives rise to a number of factor models not considered previously and enables the study of latent variables in an integrated methodological framework, rather than as a collection of seemingly unrelated special cases. The framework accommodates a great variety of different measurement scales and accommodates cases where different latent variables have different distributions. The models are estimated with the method of simulated likelihood, which allows for higher dimensional factor solutions to be estimated than heretofore. The models are illustrated on synthetic data. We investigate their performance when the distribution of the latent variables is mis-specified and when part of the observations are missing. We study the properties of the simulation estimators relative to maximum likelihood estimation with numerical integration. We provide an empirical application to the analysis of attitudes.

Key words: factor model, simulated likelihood, latent variable model.

1. Introduction

Factor models are central in psychometrics (Mulaik 1972). Maximum likelihood factor analysis was originally developed for continuous, normally distributed, observed variables, and later for binary variables (Bartholomew & Knott, 1999). Those models, however, all assume the latent variables to be normal. Our purpose is to add to the existing literature by developing a general factor model for (mixed) outcome variables in the exponential family and (mixed) normal and nonnormal latent variables. It accommodates a great variety of data, including rating, ordering, choice, frequency, and timing data and entails a number of special cases of factor analysis not considered previously. Thus, the proposed models enable the study of latent variable problems in psychology and related disciplines in an integrated methodological framework, rather than as a collection of seemingly independent special cases.

Bartholomew and Knott (1999) provide a framework for latent variable models. Their classification is based on the metrics of the observed and latent variables. Both are considered to be either discrete or continuous, leading to the classification shown in Table 1. The framework for factor models by Bartholomew and Knott integrates a collection of seemingly unrelated models

TABLE 1.
Bartholomew's classification of latent variable models

Observed Variables Latent Variables	Continuous	Discrete
Continuous	Factor analysis	Binary Factor analysis
Discrete	Finite mixtures	Latent Class Models

Requests for reprints should be sent to: Michel Wedel, Faculty of Economics, University of Groningen, PO Box 800, 9700 AV Groningen, THE NETHERLANDS. E-Mail: M.Wedel@eco.rug.nl, or Wedel@umich.edu

0033-3123/2001-4/1999-0738-A $00.75/0

for latent variable analysis and forms the basis of the models we develop. We first establish some notation. Denote the outcome variables by Y and the latent variables by X. A latent variable model involves a specification of $f(X)$, the distribution of the latent variables, and $f(Y|X)$, the distribution of the observed variables, given the latent variables. In estimating a factor model, the purpose is to retrieve the P latent variables on the basis of the observed variables. This problem cannot be easily solved in the general case of arbitrary distributions of the latent and observed variables, but Bartholomew and Knott show that sufficient statistics exist if the J observed variables follow a distribution in the exponential family. The exponential family is a broad class of distributions that has wide applicability, since it encompasses a variety of distributions for continuous variables, such as the normal, the exponential, the Erlang and other gamma distributions, as well as distributions for discrete variables, such as the Poisson and the binomial.

In factor analysis of psychometric data the observed variables are commonly assumed to be of a single type. However, latent-variable models are sometimes applied to problems in which variables are measured on different scales. For example, some of the variables may be measured on binary scales, while other scales have rank order and/or interval properties. In order to accommodate the different measurement scales for the observed variables, one needs to specify different distributions for different variables. Such models for multivariate data with mixed outcomes have recently attracted attention in the statistical literature. Examples are the work of Sammel, Ryan and Legler (1997), Moustaki (1996), Muthén (1984), Arminger and Kusters (1988), and Moustaki and Knott (2000). Moreover, maximum-likelihood factor models to date are based on the assumption that the factor scores follow a normal distribution. Whereas that assumption may be quite reasonable in many applications, there may also be applications where it is not. It is made merely for convenience and is in many cases not based on theory on the underlying psychological process. In some applications, for example, a skewed distribution of latent variables may be more appropriate than a symmetric one. Bartholomew and Knott (1999) provide arguments based on the central limit theorem that the effect of those distributional assumptions is largest when the number of factors is small. Bartholomew (1988) and Seong (1990) investigate the sensitivity of one-factor latent trait models to the specification of the prior distribution of the latent variables empirically, and find that the effect is negligible. However, the lack of theory makes the appropriate distribution of the factor scores an empirical issue that needs to be investigated in each application. Thus, there is a need for models that accommodate a wider class of (mixed) distributions of the latent variables to enable tests of the assumptions on the factor score distributions.

We provide such a class of models that allows both the outcome and latent variables to follow different distributions within the exponential family. We extend previous research in a number of ways: first, we accommodate more than two latent variables to be identified from nonnormal observed proxy variables, which is made possible by new developments in simulated likelihood estimation; second, we accommodate distributions of the latent variables beyond the normal; third, we deal with observed and latent variables of mixed distributional forms, and fourth, we deal with missing observations. The models are detailed in the next section.

2. The Exponential Family Factor Model

2.1. The Factor Models

Let $n = 1, \ldots, N$ denote subjects, $j = 1, \ldots, J$ variables and $p = 1, \ldots, P$ factors. We assume to have a two-way data matrix Y classified by subjects, n, and variables j. The observations $y_j = \left(y_{nj}\right)$, are realizations of random variables that have (conditional upon X) a distribution in the exponential family:

$$f_j(y_{nj}|\theta_{nj}, \phi_j) = \exp\left[\frac{y_{nj}\theta_{nj} - a_j(\theta_{nj})}{\phi_j} + b_j(y_{nj}, \phi_j)\right] \tag{1}$$

TABLE 2.
Some distributions in the univariate exponential family

Distribution	Notation	$f(y)$	Domain	Link-function
Discrete				
Binomial	$B(K,\pi)$	$\binom{K}{y}\pi^y(1-\pi)^{K-y}$	$[0, K]$	$\theta = \ln\left(\frac{\pi}{1-\pi}\right)$
Poisson	$P(\mu)$	$\frac{e^{-\mu}\mu^y}{y!}$	$(0,\infty)$	$\theta = \ln(\mu)$
Continuous				
Normal	$N(\mu,\sigma)$	$\frac{1}{\sqrt{2\pi\sigma}}\exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right]$	$(-\infty,\infty)$	$\theta = \mu$
Exponential	$G1(\mu)$	$\left(\frac{1}{\mu}\right)\exp\left[-\frac{y}{\mu}\right]$	$(0,\infty)$	$\theta = \mu^{-1}$
Erlang-2	$G2(\mu)$	$\frac{1}{2y}\left(\frac{2y}{\mu}\right)^2\exp\left[-\frac{2y}{\mu}\right]$	$(0,\infty)$	$\theta = \mu^{-1}$
Gamma	$G(\mu,v)$	$\frac{1}{y\Gamma(v)}\left(\frac{yv}{\mu}\right)^v\exp\left[-\frac{vy}{\mu}\right]$	$(0,\infty)$	$\theta = \mu^{-1}$

Here, θ_{nj} denotes the canonical parameter, ϕ_j a dispersion parameter that applies for certain distributions in the exponential family such as the normal, and $aj(\cdot)$ and $bj(\cdot)$ are functions depending on the particular distribution function of the variable j (see e.g., Fahrmeier & Tutz, 1993). The exponential family is a very useful family of distributions that has as special cases a number of well known distributions widely used in the analysis of psychometric data, including the normal, the Poisson, the Exponential, the Erlang-2, the Gamma and the Binomial. Table 2 provides an overview.

We specify $f_j(\cdot)$ to depend upon j, that is, allow each observed variable to have its own distribution. For example, we can model a set of variables, where some are described by a normal, and others by a binomial distribution. For each of those distributions, we use the canonical link function, $g(\cdot)$, to relate the canonical parameter to the expectation of the random variable. For example, for the normal distribution the canonical link is the identity link, for the binomial it is the logit link, and for the Poisson it is the log link function. We now specify the canonical parameter as a factor model:

$$\Theta = \lambda_0' + X\Lambda', \tag{2}$$

with $\Theta = (\theta_{nj})$, X the $(N \times P)$ matrix representing the scores on the latent variables, Λ a $(J \times P)$ matrix of fixed parameters and λ_0 a $(J \times 1)$ vector with an intercept for each observed variable. The number of latent variables, P, is unknown in most applications and needs to be identified from the data.

We assume that the latent variables, contained in the $(N \times P)$ matrix X, are independent and follow a distribution in the exponential family:

$$f_p(x_{np}|\xi_p, \varphi_p) = \exp\left[\frac{x_{np}\xi_p - c_p(\xi_p)}{\varphi_p} + d_p(x_{np}, \varphi_p)\right], \tag{3}$$

where ξ_p denotes the canonical parameter, φ_j the dispersion parameter, and $c_p(\cdot)$ and $d_p(\cdot)$ are functions depending on the particular distribution and latent variable p. We allow $f_p(\cdot)$ to depend upon p, that is, each latent variable may have its own distribution, but to remain with in the factor analysis framework, we consider only continuous distributions in the exponential family. Discrete distributions such as the binomial lead to latent class models that have received

a lot of attention in the psychometric literature and are not dealt with here. Note that, conditional upon X, $f_j(\cdot)$ is a member of the exponential family; the marginal distribution of y_j may not be tractable.

2.2. Notation and Examples

We introduce some additional notation to indicate the various special cases of the model described in (1) through (3). We represent the most common univariate members of the exponential family with symbols as shown in Table 2. The number of variables is indicated in parenthesis. We use the same notation for observed and latent variables, concatenated by a multiplication sign. For example, a standard factor analysis model with 10 observed normally distributed variables and 3 normally distributed latent factors is indicated as $N(10) \times N(3)$. To illustrate the generality of the approach we provide examples of a few special cases below.

1. If both $f_j(\cdot)$ and $f_p(\cdot)$ are normal distribution functions for all j and p (the subscripts could be dropped), the standard factor model arises where all observations and factor scores are assumed to be normally distributed. (Notation: $N(J) \times N(P)$.)
2. If $f_j(\cdot)$ is binomial for all j and the $f_p(\cdot)$ is the normal distribution function for all p latent variables (again the subscripts are redundant), a factor model for binary variables arises, in which the factor scores are assumed to be normally distributed. (Notation: $B(J) \times N(P)$.)
3. If $f_j(\cdot)$ is a different distribution function in the exponential family for different j, and f_p is normal for all p (thus the subscript p could be dropped), a factor model for mixed outcomes is obtained. (Notation: $NB(J_1, J_2) \times N(P)$, $BP(J_1, J_2) \times N(P)$, $GP(J_1, J_2) \times N(P)$, etc.)
4. If $f_j(\cdot)$ is an arbitrary distribution in (1) for any j, and $f_p(\cdot)$ is gamma for all p (thus the subscript p could be dropped), a factor model arises that accounts for skewed distributions of the underlying factor scores. Special cases are factor scores that follow an exponential (G1) or an Erlang (G2) distribution. (Notation: $N(J) \times G1(P)$, $B(J) \times G2(P)$, etc.)
5. If $f_j(\cdot)$ is a different distribution function in the exponential family for different j, and $f_p(\cdot)$ is a different distribution function in the exponential family for different p, the most general factor model arises in which outcome and latent variables have mixed distributional forms. (Notation: $NB(J_1, J_2) \times NG(P_1, P_2)$, $BP(J_1, J_2) \times NG(P_1, P_2)$, etc.)

Note that while examples 1 to 3 have been documented in the literature, examples 4 and 5 have not been previously described. We focus on the latter classes of models.

3. Estimation

3.1. Simulated Likelihood

We intend to estimate the model defined by (1) to (3) by maximizing the likelihood function. In the estimation we consider the unobserved factor scores, X, as missing data. We accommodate situations where part of the observations is missing, e.g., due to item non-response. Such situations appear to be relatively common in applications of factor analysis. Usually one uses list-wise or case-wise deletion of missing records, but Little and Rubin (1987) consider these procedures inappropriate. These authors describe ML-factor analysis where part of the observations are missing, assuming the data are missing at random (MAR). MAR implies that the missing data generation mechanism only depends upon the observed data, and not upon the missing data (Little and Rubin 1987), and is more general than MCAR, where the missing data generation mechanism is completely random. Recently, Kamakura and Wedel (2000) develop procedures to deal with missing data in factor analysis. Missing data can be easily accommodated in our estimation framework; assuming MAR, the factor model can be estimated from the observed data only, and the data can be imputed in a second and independent step. We denote the missing data

as Y^m, the observed data as Y^o and collect all parameters of the model in B. The observed data likelihood is obtained by integrating over the distribution of the unobserved factor scores and missing observations:

$$L(B|Y^o) = \iint f(Y^m|Y^o, X, \Lambda, \phi) f(Y^o|X, \Lambda, \phi) f(X|\xi, \varphi)\, dX\, dY^m. \tag{4}$$

The estimation of B is often not feasible, given the high-dimensional integration involved in the likelihood, since that in general cannot be evaluated numerically. Note that previous work, for example Moustaki and Knott (2000), was limited to factor models with only two factors because of the limitations of numerical integration. However, advances in simulated likelihood (SML) estimation have made the approximation of such integrals possible (Gouriéroux & Monfort 1997). Such simulation methods to solve estimation problems involving high-dimensional integration were introduced by McFadden (1989), and excellent overview is provided by Stern (1997). For the ease of notation, we first assume complete data, so that the log likelihood is

$$l(B|Y) = \sum_n \ln \left(\int \prod_j f_j(y_{nj}|X, \Lambda, \phi) \prod_p f_p(x_{np}|\xi, \varphi)\, dX \right). \tag{5}$$

The problem is to evaluate this log-likelihood (5) in the general case where X is a P-dimensional multivariate random variable with a known density, and y_n is a J-dimensional observation vector. The estimator obtained by maximizing (5), which is often done numerically, is consistent, efficient and asymptotically normal for a large class of models. However, if the dimensionality P is larger than three, standard numerical integration cannot be used. The idea of simulation is to draw S random variables z^s from $f(X|-)$ and use

$$\tilde{l}(B|Y) = \sum_n \ln \sum_s \prod_j \tilde{f}_j(y_{nj}|z^s; \Lambda, \phi)/S, \tag{6}$$

instead of (5). Instead of solving the integrals numerically, they are approximated through summations over $s = 1, \ldots, S$ draws from the distributions of the factor scores.

Now, $\tilde{l}(B|Y) \to l(B|Y)$ as $S \to \infty$, from the strong law of large numbers. Thus the simulated likelihood (6) is a consistent simulator of the likelihood (5). The value of B that maximizes (6) is the SML estimator. SML provides consistent estimators only if $S \to \infty$ as $N \to \infty$. This can be seen from

$$\lim_{N,S\to\infty} \sum_n \ln \sum_s \tilde{f}(y_n|z^s; B)/S = \lim_{N\to\infty} \sum_n \ln \int \tilde{f}(y_n|z; B)\, dz,$$

since the mean over s converges to the integral for $S \to \infty$. Because $\tilde{f}(\cdot)$ is a consistent simulator of $f(\cdot)$, the estimator is consistent and asymptotically equivalent to the ML estimator.

The procedure for maximizing (6) works as follows.

1. Fix a value of S. Draw z S times from the assumed factor score distribution, $f(X|\xi, \varphi)$, these S values z^s are stored and remain the same throughout the optimization procedure.
2. Generate an initial estimate for B, for example through a SVD of the data, transformed by the canonical link, $g(y_n)$. Since the (simulated) likelihood of exploratory factor models suffers from local optima, the use of such rational starting values, or multiple random starting values is recommended.
3. Compute the simulated likelihood function in (6) based on the stored values z^s, $s = 1, \ldots, S$. Maximize (6) numerically over B using a Newton type algorithm to find the SML estimator. Standard numerical optimization algorithms may be used, such as Newton Raphson, Conjugate gradients and Quasi-Newton Algorithms (we employed the latter, see e.g., Scales, 1985). For that purpose one needs the first order derivatives of the simulated log-likelihood function.

An appealing aspect of SML estimation is that the simulated log-likelihood function (6) is differentiable, so that the score vector can be computed analytically:

$$\frac{\partial \tilde{l}(B|Y)}{\partial B} = \sum_n \frac{\sum_s \partial \tilde{f}(y_n|z^s; B)/\partial B}{\sum_s \tilde{f}(y_n|z^s; B)}.$$

4. If the missing data are at least MAR conditional upon X, the likelihood factors into independent components for the observed and missing data, since

$$f(Y^m, Y^o|X, B) = f(Y^o|X, B) f(Y^m|X, B)$$

so that maximizing the complete likelihood is identical to maximizing the likelihood for the observed data only (Little & Rubin, 1987). Then T multiple imputations of the missing data can be generated, in a final and independent step of the algorithm, by drawing for n and j from the missing data distribution integrated over the asymptotic distribution of the parameter estimates,

$$\int f_j(y_{nj}^m|\hat{B}; Z) f(\hat{B})\, d\hat{B}.$$

The integration is performed by drawing (S times) from the asymptotic distribution of the estimates, $f(\hat{B})$, and averaging across these draws. In order to enable multiple imputations, the approximate covariance matrix of the estimates needs to be computed. Identifying restrictions must be imposed on the parameters, after which we compute the asymptotic covariance matrix of the parameter estimates, based on the expected information matrix.

3.2. Investigating the Properties of SML

Gouriéroux and Montfort (1997) show that the SML estimator is consistent if $\sqrt{N}/S \to 0$ as $N \to \infty$ and asymptotically equivalent to the MLE (see above). The bias is of order $1/S$. Simulation studies (Geweke, Keane & Runkle, 1999; Lee 1995, 1997), show that SML has good properties for finite values of S. We use $S = 100$ in the application below, which is sufficient for obtaining satisfactory properties of the estimates (Lee 1995, 1997). We investigate the performance of SML here relative to that of ML with numerical integration. We investigate the $P = 1$ factor model, since for this model numerical integration, that we use as a benchmark, works well. For numerical integration we use 48-point quadrature (Moustaki & Knott, 2000). We generate data with $N = 300$ observations and $J = 6$ variables. The observed variables are assumed to follow either a Binomial or a Normal distribution, and the latent variables either a Normal or an Erlang-2 distribution, so that we have four cells, for each of which we generate 100 data sets, analyzed with both ML and SML ($S = 100$). In each of the 100 replications the factor weights were randomly generated, constrained to the unit circle. We compute bias and precision of the factor weight estimates across all observed variables and 100 replications.

Table 3 shows the results. Figure 1 depicts them graphically. The bias of the SML procedure with $S = 100$ is comparable to that obtained with numerical integration and is low for both methods. Only in the case of a binomial dependent variable and Erlang distributed factor scores is the bias somewhat larger, but this occurs for both numerical and simulated integration. As one would expect, the precision of the estimates is lower for binary than for normally distributed outcome variables, for both methods. From the Table it appears that the bias is not uniformly higher for SML, and that it depends on the distribution of observed and latent variables. However, for both methods, bias seems small relative to the precision of the estimates. In general, the SML approximation produces more precise estimates, with the single exception to the case of normal observed variables with Erlang-2 latent variable. Thus, there seems to be a bias-variance tradeoff: while bias of SML estimates might be somewhat larger, their variance is lower. This small

TABLE 3.
Precision and bias: Simulation vs. quadrature approximations

Observed Variables		Latent Variable			
		Normal		Erlang2	
		SML	ML-Quad	SML	ML-Quad
Binomial	Precision	0.387	0.589	0.532	0.626
	Bias	0.020	0.028	−0.012	−0.046
Normal	Precision	0.095	0.460	0.125	0.057
	Bias	0.007	0.008	0.004	0.001

Monte Carlo study illustrates that simulated likelihood performs well, and, depending on the number of draws used, may be comparable to numerical integration with respect to bias and tends to provide estimates with lower variance.

3.3. Identification

The standard factor model suffers from location, scale and rotation indeterminacy. Since these do not necessarily hold for all special cases of the general model described above, we provide some detail on these indeterminacies (see Bartholomew & Knott, 1999).

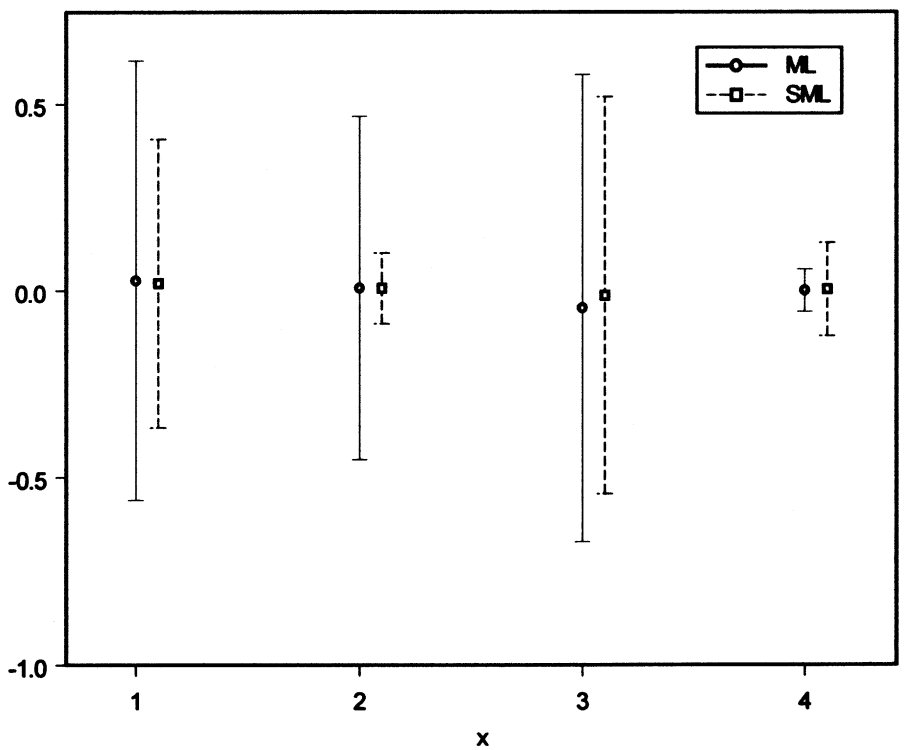

FIGURE 1.
Error bar plot of ML and SML bias and precision under four conditions: $1 = B - N$, $2 = N - N$, $3 = B - G2$, $4 = N - G2$.

A. Location and scale invariance. We have $f(Y|\lambda_0 + X\Lambda', \phi)$, which assuming $\lambda_0^* = \lambda_0 + m$, and $\Lambda^* = \Lambda C$ with $C = \text{diag}(c_p)$ a diagonal matrix, equals $f(Y|\lambda_0^* + X^*\Lambda^{*\prime}, \phi)$, where

$$X^* = (X - m(\Lambda\Lambda')^{-1}\Lambda)C^{-1}.$$

The distribution of X^* is

$$f(X^*) = \frac{1}{|C|} f_X(X^*C + m(\Lambda\Lambda')^{-1}\Lambda),$$

since the Jacobian is $|C|$, the determinant of C. The unconditional distribution of Y, integrated over the distribution of X^* is in general not equal to that integrated over the distribution of X, for members of the exponential family, with exception of, for example, the normal.

B. Rotation invariance. We have that $f(Y|\lambda_0 + X\Lambda', \phi)$, assuming $\Lambda^* = \Lambda R$ with R an orthogonal rotation matrix, equals $f(Y|\lambda_0 + X^*\Lambda^{*'}, \phi)$. Here $X^* = XR^{-1}$. The distribution of X^* is $f(X^*) = f_X(X^*R)$, since the Jacobian is 1. The unconditional distribution of Y, obtained by integrating out the distribution of X^*, is in general not equal to that obtained by integrating out that of X, for members of the exponential family, with exception of the normal distribution (Bartholomew & Knott, 1999). Thus, with the exception of those cases where the factor scores are normally distributed, the model is not rotation invariant. Rotation invariance may or may not be considered a useful property of exploratory factor models.

3.4. Inference and Model Selection

Anderson and Rubin (1956) and Gill (1977) have investigated consistency and asymptotic normality of ML estimators in factor analysis. For the models considered here these properties also follow from likelihood theory, since the standard regularity conditions hold for the general factor model in (1) to (3). The SML estimators have an asymptotic normal distribution: $\hat{B}_{SML} \sim N(B, H^{-1}UH^{-1})$, where H is the observed information matrix of second derivatives of the simulated log likelihood function, and U is the expected information matrix of the cross product of the first derivatives (Gouriéroux & Monfort, 1997). Statistical inference with Likelihood ratio and Wald tests based on the simulated likelihood function is discussed by Lee (1999). He shows that the simulated score vector can be asymptotically biased, so that limiting distributions of the simulated test statistics may be non central χ^2.

However, much of the statistical inference on parameter estimates and nested model tests are not of primary interest in evaluating exploratory factor models. For these models, inference focuses on choosing the appropriate number of factors P. Models with different numbers of factors cannot be compared using standard likelihood-based tests, since the asymptotic χ^2 distribution of the LR test of the P-factor model versus the $P + 1$-factor model does not hold (Anderson, 1980). An often used procedure is to test the $P < J$ factor model against the $P = J$ model. However, at commonly used critical values this LR test is oversensitive to small departures from the null-hypothesis due to the large number of degrees of freedom (Akaike, 1987). Aikaike developed his information criterion precisely for this model selection problem. However, the AIC statistic does not asymptotically indicate the true model among a set of candidate models. Therefore, several authors have proposed dimension-consistent criteria, such as the CAIC statistic by Bozdogan (1987). Here we use the CAIC statistic based on the simulated log-likelihood

$$\text{CAIC} = -2\tilde{l}(\hat{B}|Y^o) + K\{\ln(N) + 1\}, \tag{7}$$

with K the effective number of parameters. CAIC is dimension consistent: it indicates the "true" model as the number of observations tends to infinity. CAIC can also be used to compare models

with different assumptions on the distributions of the latent variables; since these models are not nested, the LR test cannot be applied.[1]

4. Synthetic Data Analysis

In this section we analyze synthetic data sets to further investigate the performance of our approach. We analyze data according to a variety of new models, arising as special cases of our approach, and investigate the performance of the models under varying percentages of missing observations. The synthetic data examples pertain to models with $J = 12$ mixed discrete/continuous outcome variables, and $P = 2$ latent variables that are respectively normal and Erlang-2 distributed. All data sets are generated with $N = 300$ subjects and a total of 3600 data points. The factor weights used for generating the synthetic data are presented in Table 4. The true model has a simple structure because each variable has a large weight for one factor and a small weight for the other. The intercept was taken $\Lambda_0 = -0.5$ for all variables. A range of fractions of the data are eliminated, from 0 to 50% in steps of 5%, assuming MAR. The data are generated according to the following two models:

1. The $NB(6, 6) \times N(2)$ model with mixed dependent variables and normal distributed factor scores;
2. The $NB(6, 6) \times G2(2)$ model with mixed dependent variables and Erlang-2 distributed factor scores.

First, each of the data sets is analyzed with its proper model, where we assume the proper mixed normal and binomial distributions for the observed variables. The models are calibrated using $S = 100$ random draws. For models with a normal distribution of the latent variable the solution is rotated before computing the statistics. As a performance measure we use the root mean squared error between the actual and estimated weights, RMSE($\hat{\Lambda}$).

Table 5 shows the results. When both the true and assumed scores are normal, the lowest RMSE values are obtained. There is no tendency for the RMSE to increase with the percentage of

TABLE 4.
Characteristics of the synthetic data

P	$f_p(\cdot)$	λ_0	λ_1	λ_2
1	N	−.500	−.900	.200
2	B	−.500	−.900	.100
3	N	−.500	−.800	.100
4	B	−.500	.900	.100
5	N	−.500	.900	.100
6	B	−.500	.100	.800
7	N	−.500	.200	.900
8	B	−.500	.100	.800
9	N	−.500	.200	.900
10	B	−.500	.100	−.900
11	N	−.500	.100	−.800
12	B	−.500	.100	−.900

[1] Models in which different assumptions are made on the distribution of the observed variables cannot be compared on the basis of the likelihood, since that takes on a completely different form for those different distributions. We therefore suggest an entropy-based R^2 type of statistic for comparison of those models, which is defined as (Haberman, 1982): $\rho = 1 - \ln L(Y|B)/\ln L_0(Y)$. Here $0 < \rho < 1$ can be interpreted as a multiple correlation, and $\ln L_0(Y)$ is the log-likelihood for a null-model that only includes a constant.

TABLE 5.
RMSE ($\hat{\Lambda}$) for $NB(6, 6)$ data[1]

True-Assumed score distribution	**Normal-Normal**	Normal-Erlang	**Erlang-Erlang**	Erlang-Normal
% missing				
0	**0.102**	0.153	**0.140**	0.356
5	**0.103**	0.186	**0.140**	0.331
10	**0.096**	0.178	**0.147**	0.313
15	**0.111**	0.168	**0.169**	0.334
20	**0.114**	0.142	**0.196**	0.354
25	**0.105**	0.157	**0.142**	0.328
30	**0.112**	0.180	**0.173**	0.285
35	**0.098**	0.158	**0.162**	0.323
40	**0.109**	0.156	**0.144**	0.308
45	**0.089**	0.177	**0.166**	0.329
50	**0.097**	0.179	**0.197**	0.291

[1]Results for the correctly specified model are indicated in boldface.

missing values. This indicates that missing observations do not substantially affect the recovery of the factor weights, which is partially due to the MAR assumption. The results for the Erlang model are similar, be it that the RMSE is somewhat higher than for the normal distributed factor scores. We conclude that the performance of the factor models on synthetic data is satisfactory. The RMSEs of the true and estimated factor weights indicate that the SML estimation procedure recovers the true factor weights well, in spite of the presence of missing information. The extent to which the weights are recovered appears to be affected by the mean-variance relation of the latent factor scores, and thus by the mean-variance structure of the marginal distribution of the data.

In order to investigate the effect of misspecification of the factor score distribution, we analyze the data with normal and Erlang factor scores with a misspecified factor score distribution as shown in Table 5. Table 5 reveals that the proper factor score distribution matters: the RMSEs increase substantially if the factor score distribution is misspecified. It is interesting to note that if the true scores are normal while being specified as Erlang in the model, the increase is less dramatic than when true Erlang scores are assumed to be normal in the model. This may be caused by the fact that the skewed distribution of the true factor scores in the Erlang case is not very well accommodated by the assumed normal distribution. Again, there does not seem to be an effect of the percentage of missing observations.

5. Empirical Illustration

We provide an empirical illustration of the methodology to a commercial study on consumer attitudes, which illustrates its flexibility in accommodating a variety of distributions for the observed and latent variables. We use part of the data from a commercial European study conducted by a market research agency, on attitudes toward washing clothes. We analyze a sample of 576 subjects, who rated 12 attitudinal statements on 5-point scales that are listed in Table 6. There are no missing observations in this application.

We assume rank-order properties for the measurement scales and use the binomial distribution to model them:

$$P(y_{nj} = k|B) = \binom{K-1}{k-1} p^{k-1}(1-p)^{K-k},$$

TABLE 6.
Attitude statements in laundry study

	Statement
1.	"I think men ought to help in the household"
2.	"Men should also do the washing"
3.	"My washing is always as clean as possible"
4.	"Sometimes I am afraid the washing will discolor during wash"
5.	"I think it is odd for men to do ironing"
6.	"It is important that there is fresh smell in linen-cupboard"
7.	"Washing a lot makes things wear out faster"
8.	"Washing a lot makes clothes lose color faster"
9.	"My washing always has a pleasant smell"
10.	"I usually buy new products before friends do"
11.	"I change the products I buy as little as possible"
12.	"I like to buy new and unusual things"

with p parameterized as in (2), and use the logit-link. We investigate a variety of distributions for the factor scores that allow for varying degrees of skewness: the normal distribution and two special cases of the Gamma distribution: the Erlang-2, which is more skewed than the normal, and the exponential, which is more skewed than the Erlang-2. Each of the models is estimated assuming $P = 1$ to $P = 5$ factors, to determine the optimal number needed for an adequate description of the data. $P = 4$ factors were indicated as optimal by CAIC, for all distributions of the factor scores. The $B(12) \times N(4)$ model provides the best approximation to the data, as indicated by CAIC = 19335.9, while the $B(12) \times G2(4)$(CAIC = 19408.4) does better than the $B(12) \times G1(4)$(CAIC = 19441.4). We also estimated a model with mixed latent variables: the $B(12) \times NG2(2, 2)$ model, with two factors with normal and two factors with Erlang distributed scores, to our knowledge the first application of such a model in the literature. The goodness-of-fit for this model, however, is poorer than for the model with normal factor score distributions, as indicated by the CAIC statistic (CAIC = 19430.3). In this particular application a more skewed distribution of the factor scores leads to a worse fit of the model.

We present the estimated factor weights for the four estimated 4-factor models in Table 7. Large weights are indicated in the table by boldface, which may be helpful in interpreting the latent variables. Figure 2 presents the weights of the best model, $B(12) \times N(4)$, in three dimensions, and Figure 3 provides a parallel line plot of the factor weights for the four models (after reflection of the weights for some factors). Across the four models, the first factor shows large weights of two items, 10 and 12, and to a lesser extent of item 11, and seems to capture innovative purchase behavior. Items 3, 6 and 9 have high weights for factor 2. This factor pertains to aspects of freshness and cleanliness. Factor 3 captures damage done to the color of clothes by washing, as indicated by large weights of items 4, 7 and 8. Finally, Factor 4 has to do with gender roles in washing, as indicated by large weights of items 1, 2 and 5. It may be observed from Table 7 and Figure 3 that the estimated factor weights are quite similar for the different assumed factor score distributions.

6. Conclusions

We have extended existing procedures for factor analysis to a more general class, accommodating mixed distributions of the observed and latent variables. We focus on continuous factor scores, but discrete distributions can also be dealt with (leading to finite mixture models). We develop factor models for observed variables that have a distribution in the exponential family, with more than two or three latent factors, with factor scores that have other than the normal distribution, and with factor scores that have different distribution across the latent factors. We

TABLE 7.
$P = 4$ Factor weights for various factor models estimated to the laundry data

	$B(12) \times N(4)$				$B(12) \times G1(4)$				$B(12) \times G2(4)$				$B(12) \times NG2(2, 2)$			
Item	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	−0,01	−0,02	−0,08	**−0,90**	0,05	−0,04	−0,14	**0,98**	0,03	0,11	−0,19	**0,93**	−0,11	0,12	−0,07	**0,85**
2	−0,06	−0,31	−0,25	**−1,68**	−0,12	−0,45	−0,16	**1,60**	−0,23	0,31	−0,23	**1,70**	−0,28	0,35	−0,09	**1,51**
3	−0,01	**0,76**	0,01	0,17	−0,02	**0,89**	−0,10	−0,20	−0,02	**−0,68**	0,00	−0,24	−0,11	**−0,78**	0,01	−0,24
4	−0,41	−0,20	**−1,25**	−0,02	−0,43	−0,24	**−1,46**	−0,03	−0,51	0,15	**−1,04**	0,07	−0,27	0,06	**−1,12**	−0,04
5	−0,27	0,24	−0,15	**1,93**	−0,23	0,27	−0,26	**−2,11**	−0,06	−0,38	−0,26	**−1,92**	0,16	−0,45	−0,25	**−1,71**
6	−0,02	**1,31**	−0,08	0,11	0,01	**1,44**	−0,21	−0,21	−0,10	**−1,28**	−0,01	−0,15	−0,24	**−1,37**	−0,03	−0,27
7	0,09	0,00	**−0,51**	−0,05	0,19	0,21	**−0,49**	0,11	0,23	0,05	**−0,62**	0,10	0,11	0,04	**−0,34**	−0,01
8	0,05	0,14	**−0,76**	−0,03	−0,05	0,19	**−0,87**	−0,01	0,02	−0,09	**−0,89**	0,04	−0,03	−0,17	**−0,65**	0,00
9	0,04	**0,98**	−0,01	0,03	0,04	**1,16**	−0,12	−0,11	−0,02	**−0,88**	0,01	−0,07	−0,13	**−1,00**	0,02	−0,18
10	**−0,87**	0,16	−0,11	0,15	**−0,76**	0,13	−0,11	−0,12	**−0,79**	−0,21	−0,07	−0,08	**−0,41**	**−0,38**	−0,19	−0,03
11	0,37	0,14	−0,21	−0,01	0,24	0,19	−0,21	0,00	0,24	−0,06	−0,18	−0,02	0,13	0,01	−0,06	−0,05
12	**−1,23**	0,00	−0,12	−0,10	**−1,39**	−0,13	−0,09	0,09	**−1,20**	0,07	−0,08	0,14	**−0,63**	−0,37	−0,23	0,25

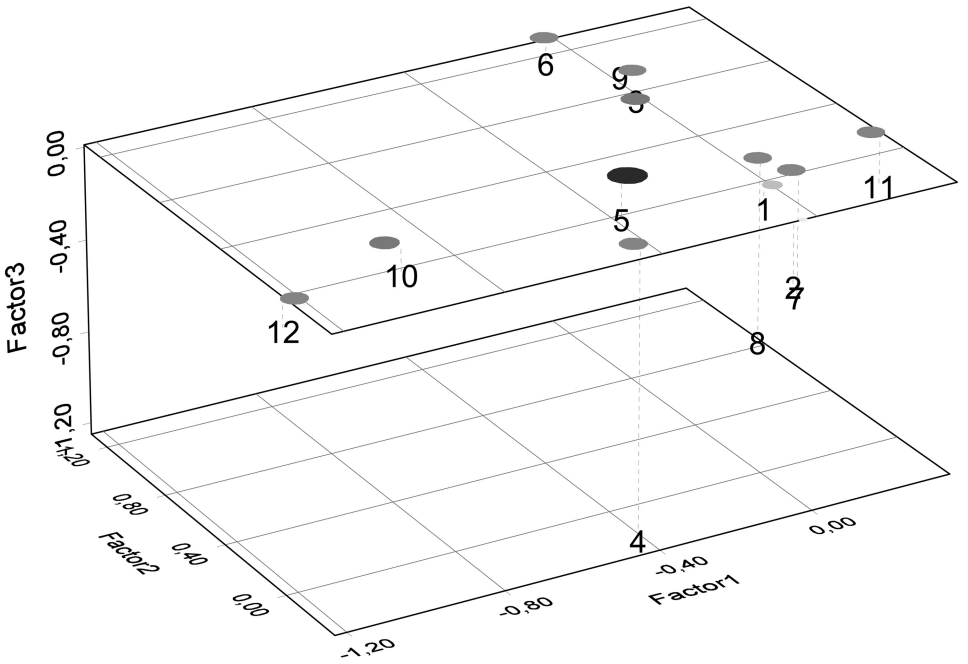

FIGURE 2.
3D scatter projection plot of factor weights of the 12 laundry attitudes. Bubble size and intensity indicate the magnitude of the weights on Factor 4.

find, however, that although simulation studies show that the recovery of the factor weights deteriorates if the factor score distribution is misspecified, the estimated factor weights themselves are rather invariant under various assumed factor score distributions in our empirical application. Previously, such robustness was found (for one-factor latent trait models), for example, by Bartholomew (1988) and Seong (1990), and our results corroborate their findings. Whether that holds more general than in our application to one single data set needs to be investigated in future empirical studies. However, an advantage of nonnormal factor scores is that these alleviate the rotational indeterminacy of traditional factor models, and thus eliminate the need to apply rotation methods for interpretation. But, rotational invariance of factor models may or may not be considered an asset: invariance allows for rotations that may be helpful in interpretation of factor solutions, but on the other hand renders models underidentified so that standard errors cannot be computed without imposing further constraints.

The estimation of high dimensional factor models is made possible by developments in simulated maximum likelihood estimation. We show in analysis of synthetic data that the SML estimators (with $S = 100$) have comparable bias and smaller variance than the standard ML approach with numerical integration. We investigate the performance of the SML approach for one and two factor models and found it to be quite satisfactory. The performance of the simulation method, however, depends on the dimension of the integration involved (Lee, 1995, 1997). For higher dimensional models, more efficient sampling methods have been developed, for example methods based on antithetic acceleration (Geweke, 1988), Latin Hypercubes (McKay, Conover, & Beckman, 1979) and Orthogonal Arrays (Owen, 1992; Tang, 1993). See for an overview and comparison of these methods Sándor and András (2000). Those sampling methods are more efficient and deserve further investigation in the context of the estimation of factor models. However, other methods for approximating integrals through simulation are also currently available, in particular in the area of Bayesian computation (Markov Chain Monte Carlo Methods; see Ansari & Jedidi, 2000, for an application to factor analysis for binary data). While we cannot conclude

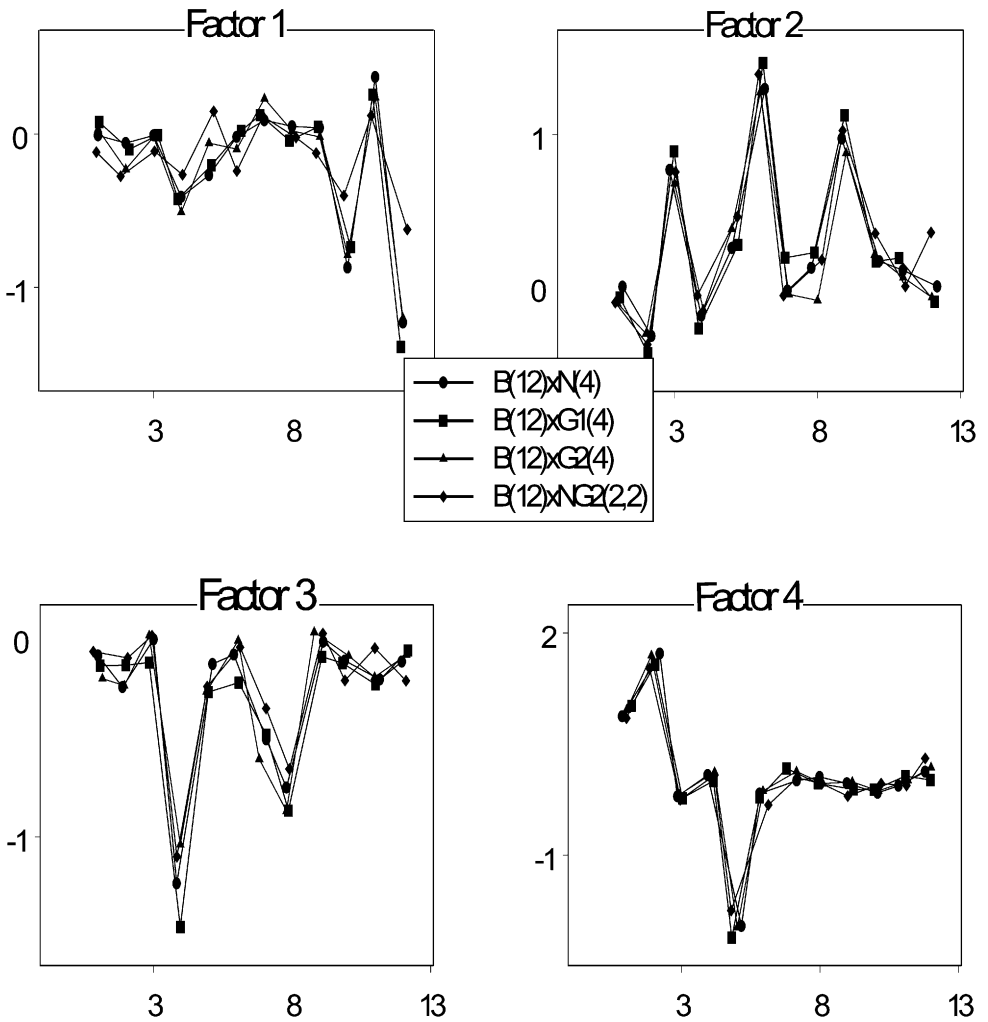

FIGURE 3.
Line plots of the weights of the 12 items (X axis) on the four factors.

SML to be universally superior, an extensive comparison being beyond the scope of the present paper, we would like to mention the advantages of SML over MCMC of not being dependent upon the specification of (informative or uninformative) prior distributions of the parameters, and of enabling the application of standard numerical optimization routines, so that assessment of convergence is easy.

There are a number of limitations that need mentioning. First, the approach requires the specification of the distribution of the latent variables. Currently, there are no methods for checking the accuracy of those distribution assumptions before the model is fitted. The distribution of the factor scores needs to be decided upon by specifying several alternative distributions, estimating the models and comparing their fit, as in our empirical application. Secondly, our procedure currently does not accommodate covariance between the latent variables. However, in the case where the factor score distribution is normal, it is simple to obtain the draws from a multivariate distribution. Another limitation is that model search for models with mixed latent variables, is highly involved due to the exploding number of models. For example, if $P = 1$ to $P = 6$ factors are considered and there are two factor score distributions 27 models need to be considered. We therefore employed the heuristic strategy in the application to first search for the optimal value

of P based on a single distribution of the latent variables and then fit a mixed model conditional upon the value of P. The same model selection problem applies to models with mixed discrete and continuous latent variables. The present study was restricted to continuous latent variables. Including discrete and continuous latent variables leads to potentially interesting new classes of models that need to be further explored. Although these models are accommodated in our modeling framework, that is based on Bartholomew and Knott (1999), further research is needed into the properties and applicability of these models and into issues of selecting the appropriate number of mixed continuous latent variables and latent classes. A final limitation of the approach is that it takes more computer resources than standard factor analysis, the typical estimation run taking several hours, due to the simulated likelihood procedure. However, we expect this to become less and less of a problem with the rapid increase in the speed and internal memory of desktop computers.

References

Akaike H., (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

Anderson, E.B. (1980). *Discrete statistical models with social science applications*. New York, NY: North Holland.

Anderson, T.W., & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the third Berkeley Symposium in Mathematical Statistics and Probability*, *5*, 111–150.

Ansari, A., & K. Jedidi (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika*, 475–496.

Arminger, G., & Kusters, U. (1988). Latent trait models with indicators of mixed measurement level, In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 51–71). New York, NY: Plenum.

Bartholomew, D.J. (1988). The sensitivity of latent trait analysis to choice of prior distribution. *British Journal of Mathematical and Statistical Psychology*, *41*, 101–107.

Bartholomew, D.J., & Knott, M. (1999). *Latent variable models and factor analysis* (Kendalls Library of Statistics, No. 7, 2nd. ed.). New York, NY: Edward Arnold.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345–370.

Fahrmeier, L., & Tutz, G. (1991). *Multivariate statistical modeling based on generalized linear models*. New York, NY: Springer-Verlag.

Geweke, J.M. (1988). Anthitetic acceleration of Monte Carlo integration in Bayesian inference. *Journal of Econometrics*, *57*, 1317–1339.

Geweke, J., Keane, M., & Runkle, D. (1994). Alternative computational approaches to inference in the multinomial probit model. *Review of Economics and Statistics*, *76*(4), 609–632.

Gill, R.D. (1977). Consistency of maximum likelihood estimator of the factor analysis model when the observations are not multivariate normal. In J.R. Bara, F. Brodeau, G. Romier, & B. van Cutsem (Eds.), *Recent developments in statistics* (pp. 437–440). Amsterdam: North Holland.

Gouriéroux, C., & Monfort, A. (1997). *Simulation based econometric methods*. New York, NY: Oxford University Press.

Haberman, S.J. (1982). Analysis of dispersion of multinomial responses. *Journal of the American Statistical Association*, *77*, 568–580.

Kamakura, W.A., & Wedel, M. (2000). Factor analysis and missing data. *Journal of Marketing Research*. *37*, 490–498.

Lee, L.F. (1995), Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory*, *11*, 437–483.

Lee, L.F. (1997). Simulated maximum likelihood estimation of dynamic discrete choice statistical models: Some Monte Carlo results. *Journal of Econometrics*, *82*, 1–35.

Lee, L.F. (1999). Statistical inference with simulated likelihood functions. *Econometric Theory*, *15*, 337–351.

Little, R.J.A., & D.B. Rubin (1987). *Statistical analysis with missing data*. New York, NY: Wiley.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrika*, *57*, 995–1026.

McKay, M.D., Conover, W.J. & Beckman, R.J. (1979). A Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*, 239–245.

Moustaki, I. (1996). A latent trait and latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology*, *49*, 313–334.

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

Mulaik, S.A. (1972). *The foundations of factor analysis*. New York, NY: McGraw Hill.

Muthén, B.O. (1984). A general structural model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.

Owen, A. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, *2*, 439–452.

Sammel, M.D., Ryan, L.M., & Legler, J.M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society*, *Series B*, *59*(3), 667–678.

Sándor, Z., & András, P. (2000). *Alternative sampling methods for estimating multivariate normal probabilities* (Working Paper). Groningen, Netherlands: University of Groningenm Faculty of Economics.

Scales, L.E. (1985). *Introduction to non-linear optimization*. London: Macmillan.

Seong, T.J. (1990). Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement*, *14*, 299–311.
Stern, S. (1997). Simulation-based estimation. *Journal of Economic Literature*, *35*(December), 2006–2039.
Tang, B. (1993). Orthogonal array-based hypercubes. *Journal of the American Statistical Association*, *88*, 1392–1397.

Manuscript received 18 MAY 1999
Final version received 4 DEC 2000