Journal of Educational and Behavioral Statistics
Spring 2001, Vol. 26, No. 1, pp. 51–71

Multidimensional Item Response Theory Modeling of Binary Data: Large Sample Properties of NOHARM Estimates

Albert Maydeu-Olivares
Universidad Carlos III de Madrid

Keywords: *multidimensional normal ogive model, categorical data analysis, asymptotic theory, limited information*

NOHARM is a program that performs factor analysis for dichotomous variables assuming that these arise from an underlying multinormal distribution. Parameter estimates are obtained by minimizing an unweighted least squares function of the first- and second-order marginal proportions. Here, large sample standard errors for restricted as well as rotated unrestricted factor solutions are given. Also a test of the goodness of fit of the model to the first- and second-order marginals of the contingency table is proposed. In a simulation study, it was found that for small models, accurate parameter estimates, standard errors, and goodness-of-fit tests can be obtained with as few as 100 observations. Furthermore, NOHARM estimates, standard errors, and goodness-of-fit tests are comparable to those obtained using a related LISREL procedure.

NOHARM (Normal Ogive Harmonic Analysis Robust Method; Fraser & McDonald, 1988) is a program that estimates the common factor model from a set of n binary variables assuming that these arise by dichotomizing a n-variate normal density. In the educational testing context, this model is generally referred to as multidimensional normal ogive (MNO) model. The program is based on theory for nonlinear factor analysis developed by McDonald (1967, 1982a, 1982b, 1985) and it is one of the most widely used programs for multidimensional item response theory (IRT) modeling of binary educational data (Reckase, 1997). An up to date presentation of the underlying theory is given in McDonald (1997).

The NOHARM program is characterized by two features:

1) It fits an approximation to the MNO model rather than the MNO model itself.
2) This approximating model is estimated from the first- and second-order marginals of the contingency table using a two-stage least squares approach.

This research was supported by Grant PM97-0113 from the DGESIC, Spanish Ministry of Education and Science. The author is indebted to Michael Browne, two anonymous reviewers, and very specially, Rod McDonald, for helpful comments to a previous draft.

Three additional features make this program particularly attractive. First, it is able to estimate very large models and the estimation is extremely fast. Second, it estimates restricted as well as unrestricted factor models, where the latter can be rotated to approximate simple structure by Varimax (Kaiser, 1958) or Promax (Hendrickson & White, 1964) rotation methods. Third, simulation studies by Knol and Berger (1991) shows that the program provides comparable (if not better) parameter estimates than full information maximum likelihood (see Bock & Aitkin, 1981).

The major drawback of NOHARM is that no standard errors for its parameter estimates nor tests of the goodness of fit of the model have been available (McDonald, 1997). This article fills this gap using large sample theory. Here, we first describe the MNO model and an approximation to this model (McDonald, 1967, 1982b) which is used in NOHARM. Next, we describe the two-stage limited information estimation procedure implemented in the NOHARM program and two closely related estimators, the one-stage estimation procedure of Christoffersson (1975) and the three-stage estimation procedure proposed by Muthén, (1978, 1993; Muthén, du Toit, & Spisic, in press). Then, we review relevant statistical theory for weighted least squares estimation of moment structures and show how this can be applied to obtain asymptotic standard errors and goodness-of-fit tests for NOHARM estimates. To illustrate our results, we provide the NOHARM standard errors and goodness-of-fit test for a one-dimensional restricted model, a two-dimensional restricted model, and a two-dimensional unrestricted model with Varimax rotation for the widely studied LSAT 7 data (Bock & Lieberman, 1970). Finally, we perform a small simulation study to investigate the small sample performance of the proposed asymptotic standard errors and goodness-of-fit tests, and to compare the behavior of the estimator implemented in NOHARM with a two-stage estimator in which the NOHARM approximation to the MNO model is not used, and to Muthén's (1993) estimator. The latter is available in LISREL 8 (Jöreskog & Sörbom, 1993).

The Multidimensional Normal Ogive Model

Let **y*** be an n-dimensional normal density with a p-dimensional factor structure

$$\mathbf{y}^* \sim N(\mathbf{0}, \mathbf{\Lambda\Psi\Lambda}' + \mathbf{\Theta}), \tag{1}$$

where $\mathbf{\Lambda}$ is an $n \times p$ matrix of factor loadings, $\mathbf{\Psi}$ is a $p \times p$ matrix of interfactor correlations, and $\mathbf{\Theta}$ denotes an $n \times n$ diagonal matrix of residual variances (uniquenesses). Assume that each y_i^* has been dichotomized by a threshold τ_i as

$$y_i = \begin{cases} 1 & \text{if} \quad y_i^* \geq \tau_i \\ 0 & \text{if} \quad y_i^* < \tau_i. \end{cases} \tag{2}$$

It then follows that

$$\Pr\left[\bigcap_{i=1}^{n} y_i\right] = \int \cdots_{\mathbf{R}} \int \phi_n(\mathbf{y}^*)\, d\mathbf{y}^* \qquad i = 1, \ldots, n, \tag{3}$$

where $\phi_n(\cdot)$ denotes an n-dimensional standard normal density and $\mathbf{R}$ is an n-dimensional area of integration with intervals

$$R_i = \begin{cases} (\tau_i, \infty) & \text{if} \quad y_i = 1 \\ (-\infty, \tau_i) & \text{if} \quad y_i = 0. \end{cases} \tag{4}$$

As the variances of **y*** are not identifiable from binary data, to identify the model we let

$$\mathbf{\Theta} = \mathbf{I} - diag\,(\mathbf{\Lambda\Psi\Lambda}'), \tag{5}$$

so that **y*** has unit variances.

In particular,

$$\pi_i = \Pr(y_i = 1) = \int_{\tau_i}^{\infty} \phi_1\left(y_i^*\right) dy_i^* = \Phi_1(-\tau_i) \qquad i = 1, \ldots, n \tag{6}$$

$$\pi_{ij} = \Pr(y_i = 1, y_j = 1) = \int_{\tau_i}^{\infty}\int_{\tau_j}^{\infty} \phi_2\left(y_i^*, y_j^*: \rho_{ij}\right) dy_i^* dy_j^* = \Phi_2(-\tau_i, -\tau_j, \rho_{ij}),$$
$$i = 2, \ldots, n;\ j = 1, \ldots, i-1 \tag{7}$$

where $\Phi_n(\cdot)$ denotes an n-dimensional standard normal distribution function, and $\rho_{ij} = \boldsymbol{\lambda}_i'\mathbf{\Psi}\boldsymbol{\lambda}_j$.

Equation (3) can be rewritten as (Takane & de Leeuw, 1987)

$$\Pr\left[\bigcap_{i=1}^{n} y_i\right] = \int_{\hat{R}} \phi_p(\boldsymbol{\eta}) \prod_{i=1}^{n} \Pr(y_i = 1|\boldsymbol{\eta})^{y_i}[1 - \Pr(y_i = 1|\boldsymbol{\eta})]^{1-y_i}\, d\boldsymbol{\eta} \tag{8}$$

$$\Pr(y_i = 1|\boldsymbol{\eta}) = \Phi_1\left[\frac{-\tau_i + \boldsymbol{\lambda}_i'\boldsymbol{\eta}}{\sqrt{1 - \boldsymbol{\lambda}_i'\mathbf{\Psi}\boldsymbol{\lambda}_i}}\right], \tag{9}$$

where $\hat{\mathbf{R}}$ is a p-dimensional area of integration with intervals $(-\infty, \infty)$, and $\boldsymbol{\eta}$ denotes the p-variate vector of unobserved latent traits.

In the educational testing literature, the probability of endorsing an item given the latent traits, [i.e., $\Pr(y_i = 1 \mid \boldsymbol{\eta})$], is referred to as an *item response function* (e.g., van der Linden & Hambleton, 1997). To simplify the estimation of this model, McDonald (1967, 1982b) sought to approximate the strictly nonlinear item response function (9) by a polynomial function. He showed that

$$\Phi_1\left(\frac{-\tau_i + \boldsymbol{\lambda}_i'\boldsymbol{\eta}}{\sqrt{1 - \boldsymbol{\lambda}_i'\boldsymbol{\Psi}\boldsymbol{\lambda}_i}}\right) = \Phi_1(-\tau_i) + \phi_1(\tau_i)\sum_{k=1}^{\infty} \frac{(\boldsymbol{\lambda}_i'\boldsymbol{\Psi}\boldsymbol{\lambda}_i)^{1/2}}{k!} H_{k-1}(\tau_i) H_k\left(\frac{\boldsymbol{\lambda}_i'\boldsymbol{\eta}}{\sqrt{\boldsymbol{\lambda}_i'\boldsymbol{\Psi}\boldsymbol{\lambda}_i}}\right), \tag{10}$$

where $H_k(x)$ is a Hermite polynomial of degree k satisfying

$$H_k(x)\phi(x) = (-1)^k \frac{\partial^k \phi(x)}{\partial x^k}$$

(Christoffersson, 1975, Appendix 1). This first four terms of this polynomial are

$$H_k(x) = \begin{cases} 1 & \text{if} \quad k = 0 \\ x & \text{if} \quad k = 1 \\ x^2 - 1 & \text{if} \quad k = 2 \\ x^3 - 3x & \text{if} \quad k = 3 \end{cases} \tag{11}$$

McDonald (1982b) concluded that (9) can be reasonably approximated in the interval $\boldsymbol{\eta} \in (-3, 3)$ by a third-degree polynomial. Consequently, he suggested approximating the MNO model by (8) with (10) setting $k = 3$. This approximating model is used in the NOHARM program and henceforth will be referred to as the NOHARM model.

We show in Appendix A that (8) with (10) implies that $\pi_i = \Phi_i(-\tau_i)$ and also that

$$\pi_{ij} = \Phi_i(-\tau_i)\,\Phi_i(-\tau_j) + \phi_i(-\tau_i)\,\phi_i(-\tau_j)\sum_{k=1}^{\infty} \frac{\rho_{ij}^k}{k!} H_{k-1}(\tau_i) H_{k-1}(\tau_j). \tag{12}$$

This is the well-known tetrachoric series expansion of π_{ij} (Kendall, 1941). As McDonald (1985, p. 136) noted: "the same result can be obtained either as the solution to the problem of evaluating a double integral connected with the normal distribution or as a solution to the problem of approximating a strictly nonlinear regression function by a polynomial regression function."

Limited Information Estimation

Let $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)'$ where $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ are the n-dimensional and $n(n-1)/2$ -dimensional vectors of first- and second-order marginal probabilities defined by (6) and (7), respectively, with sample counterparts $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2)'$. McDonald (1982b) suggested

estimating the NOHARM model using the following two-stage procedure: In the first stage, from (6) each threshold τ_i is estimated separately from its corresponding first-order proportion p_i as

$$\hat{\tau}_i = -\Phi_i^{-1}(p_i) \qquad i = 1, \ldots, n. \tag{13}$$

In the second stage, the first-stage parameter estimates are inserted in (7), and the distinct elements of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ collected in a q-dimensional vector $\boldsymbol{\theta}$ are estimated as the solution to

$$F_2(\boldsymbol{\theta}) = Min\,[\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})]'\,[\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})] \tag{14}$$

where each element of $\boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})$ is approximated by (12) with $k = 3$ inserting the first-stage parameters. This estimation procedure has been implemented in the program NOHARM (Fraser & McDonald, 1988). NOHARM allows the estimation of models with restrictions in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$, as well as the estimation of unrestricted models. In the case of restricted models (a.k.a. confirmatory models), it is assumed that enough restrictions have been imposed on $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ so that the model is identified. In unrestricted models (a.k.a. exploratory factor analysis) in which no prior knowledge of the structure of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ is available, the minimization when $p > 1$ is performed with respect to $\boldsymbol{\Psi} = \mathbf{I}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_L$, a low echelon matrix (i.e., $\Lambda_{L(ij)} = 0, i < j, j = 2, \ldots, p$). The resulting $\boldsymbol{\Lambda}_L$ is then rotated to approximate simple structure by means of a planar algorithm (see ten Berge, Knol, & Kiers, 1988) using (a) a Varimax criterion with Kaiser's row normalization and (b) a Promax criterion.

It is worth reviewing here two additional estimators that use the same amount of information as this two-stage estimator. One of them simply consists of estimating all model parameters in a single stage using

$$F_1(\boldsymbol{\tau}, \boldsymbol{\theta}) = Min\,[\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\tau}, \boldsymbol{\theta})]'\,\mathbf{W}[\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\tau}, \boldsymbol{\theta})]. \tag{15}$$

Christoffersson (1975) suggested employing (15) using as weight matrix the inverse of a consistent estimate of the asymptotic covariance matrix of $\mathbf{p}$. This estimator requires the numerical evaluation of all the univariate and bivariate integrals (6) and (7) at each iteration of the estimation procedure and therefore this approach is considerably slower than a two-stage estimator in which only the bivariate integrals are to be evaluated at each iteration of the estimation procedure.

Interestingly, Christoffersson (1975) also used (12) to approximate π_{ij}. However, because his rationale for using (12) was to approximate π_{ij}, he used $k = 10$. Equation (12) with $k = 3$ provides a rather poor approximation to π_{ij} and the approximation is not uniform. For instance, when $\tau_i = \tau_j = 0$, π_{ij} has a closed form solution,

$$\pi_{ij} = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho_{ij}$$

(David, 1953). In this case, Equation (12) with $k = 3$ yields an error of approximation less than 10^{-4} when $|\rho| < .5$ but it can be as high as 10^{-1} when $|\rho| = .9$. An interesting question is how the two-stage estimation procedure coupled with NOHARM's approximating model compares with a two-stage procedure in which the MNO model is fitted instead. In a review of algorithms for bivariate normal integration, Terza and Welland (1991) found Divgi's (1979) approximation to π_{ij}, which provides a uniform accuracy of the order of 10^{-7}, to "decisively outperform" competing approaches. In this paper, we shall compare the empirical behavior of the two-stage estimator when the bivariate normal integrals are approximated as in Divgi with the behavior of NOHARM, which amounts to approximating these integrals using (12) with $k = 3$.

An alternative approach to estimating the MNO model is to employ a three-stage approach. In a first stage each threshold is estimated separately from its corresponding first-order proportion p_i as in (13). In a second stage, each tetrachoric correlation ρ_{ij} is estimated separately from its corresponding bivariate proportion p_{ij} given the first-stage estimates by

$$\hat{\rho}_{ij} = \Phi_2^{-1}(p_{ij} | -\hat{\tau}_i, -\hat{\tau}_j). \tag{16}$$

Finally, if no restrictions are imposed on the thresholds $\boldsymbol{\tau}$, $\boldsymbol{\theta}$ is estimated as the solution to

$$F_3(\boldsymbol{\theta}) = Min\,[\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\theta})]'\,\mathbf{W}\,[\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\theta})] \tag{17}$$

where $\boldsymbol{\rho} = vecr(\mathbf{P} = \boldsymbol{\Lambda\Psi\Lambda}' + \boldsymbol{\Theta})$, and $vecr(\cdot)$ denotes an operator that stacks the lower diagonal elements excluding the diagonal into a column vector. Muthén has considered this three-stage estimator with a weight matrix equal to (a) a consistent estimate of the inverse of the asymptotic covariance matrix of the tetrachoric correlations (WLS: Muthén, 1978), (b) an identity matrix (ULS: Muthén, 1993), and (c) a consistent estimate of the inverse of the asymptotic variances of the tetrachoric correlations (DWLS: Muthén, et al., in press).

Since in this approach the bivariate integrals are evaluated separately for each tetrachoric correlation, this approach is computationally faster than the two-stage approach. However, it is unclear whether the three-stage estimator is computationally faster than the two-stage estimator coupled with the NOHARM model. This is investigated in this article by means of a simulation study.

Large Sample Properties of NOHARM Estimates for Restricted Factor Models

Under random sampling of the respondents, $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)'$ are the first- and second-order uncentered joint moments of a multivariate Bernoulli distribution (Teugels, 1990). Therefore, the estimation procedure implemented in NOHARM is in fact a two-stage moment structure estimator, and we may rely on existing theory for these estimators to obtain most of the desired results. Before proceeding, we will review some of the relevant theory.

Consider a random sample from a n-variate distribution. Let $\boldsymbol{\sigma}$ be a vector of joint moments of this distribution with sample counterparts $\mathbf{s}$, such that $\sqrt{N}(\mathbf{s} - \boldsymbol{\sigma}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma})$. Consider also a moment structure for $\boldsymbol{\sigma}$, $\boldsymbol{\sigma}(\boldsymbol{\vartheta})$, with Jacobian matrix $\boldsymbol{\Delta} = \partial\boldsymbol{\sigma}/\partial\boldsymbol{\theta}'$, and suppose we estimate $\boldsymbol{\vartheta}$ as the solution to

$$F(\boldsymbol{\vartheta}) = Min\,[\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\vartheta})]'\,\hat{\mathbf{W}}\,[\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\vartheta})], \tag{18}$$

where $\hat{\mathbf{W}}$ is a matrix converging in probability to $\mathbf{W}$, a non-negative definite matrix. Then, under typical regularity conditions, it follows that (e.g., Browne, 1984; Satorra, 1989; Satorra & Bentler, 1994)

$$\sqrt{N}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \overset{a}{=} (\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W}\sqrt{N}(\mathbf{s} - \boldsymbol{\sigma}) \tag{19}$$

$$\sqrt{N}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \xrightarrow{d} N[\mathbf{0}, (\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Gamma}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta})^{-1}] \tag{20}$$

$$\sqrt{N}[\mathbf{s} - \boldsymbol{\sigma}(\hat{\boldsymbol{\vartheta}})] \xrightarrow{d} N\left\{\mathbf{0}, [\mathbf{I} - \boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W}]\,\boldsymbol{\Gamma}\,[\mathbf{I} - \boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W}]'\right\} \tag{21}$$

$$N\hat{F} \xrightarrow{d} \sum_{i=1}^{r} \alpha_i \chi_1^2, \tag{22}$$

where $\xrightarrow{d}$ denotes convergence in distribution and $\overset{a}{=}$ denotes asymptotic equality. In (22) then χ_1^2's are independent chi-square variables with one degree of freedom and the α_i's are the non-null eigenvalues of

$$\mathbf{W}[\mathbf{I} - \boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W}]\boldsymbol{\Gamma}. \tag{23}$$

To apply these results to our problem, we first note that

$$\sqrt{N}[\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\tau}, \boldsymbol{\theta})] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}), \tag{24}$$

where $\boldsymbol{\Gamma}$ depends on joint moments of the multivariate Bernoulli up to order four. Christoffersson (1975, Appendix 2) provides explicit expressions for the elements of $\boldsymbol{\Gamma}$.

Let $\boldsymbol{\Delta}_{11} = \frac{\partial \boldsymbol{\pi}_1}{\partial \boldsymbol{\tau}'}$, $\boldsymbol{\Delta}_{21} = \frac{\partial \boldsymbol{\pi}_2}{\partial \boldsymbol{\tau}'}$, $\boldsymbol{\Delta}_\rho = \frac{\partial \boldsymbol{\pi}_2}{\partial \boldsymbol{\rho}'}$, $\boldsymbol{\Delta}_\theta = \frac{\partial \boldsymbol{\rho}}{\partial \boldsymbol{\theta}'}$ so that $\boldsymbol{\Delta}_{22} = \frac{\partial \boldsymbol{\pi}_2}{\partial \boldsymbol{\theta}'} = \boldsymbol{\Delta}_\rho \boldsymbol{\Delta}_\theta$. $\boldsymbol{\Delta}_{11}$ is a diagonal matrix with elements $-\phi_i(\tau_i)$. $\boldsymbol{\Delta}_{21}$ has elements

$$\frac{\partial \pi_{ij}}{\partial \tau_r} = \begin{cases} -\phi_1(\tau_i)\Phi_1\left(\dfrac{-\tau_j + \rho_{ij}\tau_i}{\sqrt{1-\rho_{ij}^2}}\right) & \text{if } r = 1 \\ -\phi_1(\tau_j)\Phi_1\left(\dfrac{-\tau_i + \rho_{ij}\tau_j}{\sqrt{1-\rho_{ij}^2}}\right) & \text{if } r = j \\ 0, \quad \text{otherwise} \end{cases} \tag{25}$$

and $\boldsymbol{\Delta}_\rho$ is a diagonal matrix with elements $\phi_2(\tau_i, \tau_i : \rho_{ij})$, a bivariate standard normal density function with parameter ρ_{ij} (Muthén, 1978).

Now, NOHARM'S first-stage estimates $\hat{\boldsymbol{\tau}}$ can be expressed as the solution to

$$F_1(\boldsymbol{\tau}) = Min[\mathbf{p}_1 - \boldsymbol{\pi}_1(\boldsymbol{\tau})]'[\mathbf{p}_1 - \boldsymbol{\pi}_1(\boldsymbol{\tau})] \tag{26}$$

so that (26) is a special case of (18) in which $\mathbf{W} = \mathbf{I}$. We note that since the relation between $\mathbf{p}_1$ and $\hat{\boldsymbol{\tau}}$ is one-to-one, $\hat{F}_1 \equiv 0$. Now, $(\boldsymbol{\Delta}_{11}'\boldsymbol{\Delta}_{11})^{-1}\boldsymbol{\Delta}_{11}' = \boldsymbol{\Delta}_{11}^{-1}$. Using this, it follows from (20) that the asymptotic covariance matrix of the thresholds estimated by NOHARM is

$$Acov(\hat{\boldsymbol{\tau}}) = \frac{1}{N}\boldsymbol{\Delta}_{11}^{-1}\boldsymbol{\Gamma}_{11}\boldsymbol{\Delta}_{11}^{-1}, \tag{27}$$

where $\boldsymbol{\Gamma}_{11}$ the asymptotic covariance matrix of $\sqrt{N}\mathbf{p}_1$ (i.e., the $n \times n$ upper left submatrix of $\boldsymbol{\Gamma}$) Alternatively, (27) follows directly from the fact that

$$\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\pi}_1'} = \left(\frac{\partial \boldsymbol{\pi}_1}{\partial \boldsymbol{\tau}'}\right)^{-1} = \boldsymbol{\Delta}_{11}^{-1}$$

(Muthén, 1978) and the multivariate delta theorem (Rao, 1973).

NOHARM's second-stage estimator is also a weighted least squares estimator similar to (18). However, to be able to apply (20) we need the asymptotic distribution of the sample statistics used in NOHARM's second stage. In Appendix B we show that

$$\sqrt{N}[\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}\,|\,\hat{\boldsymbol{\tau}})] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}) \qquad \boldsymbol{\Omega} = \left(-\boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}^{-1}\middle|\,\mathbf{I}_p\right)\boldsymbol{\Gamma}\left(-\boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}^{-1}\middle|\,\mathbf{I}_p\right)'. \tag{28}$$

Hence combining (28) with (20) we obtain

$$Ac\hat{o}v(\hat{\boldsymbol{\theta}}) = \frac{1}{N}\left(\hat{\boldsymbol{\Delta}}_{22}'\hat{\boldsymbol{\Delta}}_{22}\right)^{-1}\hat{\boldsymbol{\Delta}}_{22}'\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\Delta}}_{22}\left(\hat{\boldsymbol{\Delta}}_{22}'\hat{\boldsymbol{\Delta}}_{22}\right)^{-1}, \tag{29}$$

where $\hat{\boldsymbol{\Delta}}_{22}$ is $\boldsymbol{\Delta}_{22}$ evaluated at $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}})$ and $\hat{\boldsymbol{\Omega}}$ is obtained by also evaluating $\boldsymbol{\Delta}_{11}$, $\boldsymbol{\Delta}_{21}$ at $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}})$ and by substituting sample proportions for probabilities to consistently

estimate $\boldsymbol{\Gamma}$. Note that because of (5), the diagonal elements of $\boldsymbol{\Theta}$ are not free parameters. Instead they depend on $\boldsymbol{\theta}$. Standard errors for these parameters can be obtained from (29) by the multivariate delta method.

Consider now the residual second-order marginal proportions $\hat{\boldsymbol{\epsilon}} := [\mathbf{p}_2 - \boldsymbol{\pi}_2(\hat{\boldsymbol{\theta}}\,|\,\hat{\boldsymbol{\tau}})]$. Akin to (21) we find that

$$\sqrt{N}\hat{\boldsymbol{\epsilon}} \xrightarrow{d} N(\mathbf{0}, \mathbf{H}\boldsymbol{\Omega}\mathbf{H}')/N \tag{30}$$

where

$$\mathbf{H} = \mathbf{I}_{\frac{n(n-1)}{2}} - \boldsymbol{\Delta}_{22}(\boldsymbol{\Delta}_{22}'\boldsymbol{\Delta}_{22})^{-1}\boldsymbol{\Delta}_2'.$$

Dividing each residual $\hat{\epsilon}_{ij}$ by its standard error, where

$$\mathrm{SE}(\hat{\boldsymbol{\epsilon}}) = \sqrt{vecdiag(\mathbf{H}\boldsymbol{\Omega}\mathbf{H}')/N},$$

we obtain an adjusted residual, $\hat{a}_{ij}$, which is asymptotically standard normal. These adjusted statistics can be useful diagnostics for assessing the misfit of the model.

Now, consider the test statistic $T = N\hat{F}_2 = N\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}$. Analogous to (22), T is asymptotically distributed as a weighted sum of independent chi-square distributions with one degree of freedom each, where the weights are the eigenvalues of $\mathbf{H}\boldsymbol{\Omega}$. To assess the goodness of fit of the model to the first- and second-order marginals of the contingency table, we simply propose scaling T by its mean or adjusting it by its mean and variance as follows:

$$T_s = \frac{r}{Tr(\mathbf{H}\boldsymbol{\Omega})}T \qquad T_s = \frac{Tr(\mathbf{H}\boldsymbol{\Omega})}{Tr\left((\mathbf{H}\boldsymbol{\Omega})^2\right)}T, \tag{31}$$

where T_s and T_a denote the scaled (for mean) and adjusted (for mean and variance) test statistics, and $r = [n(n-1)/2] - q$ is the number of degrees of freedom of the model. Satorra and Bentler (1994) suggested referring T_s to a chi-square distribution with r degrees of freedom, and referring T_a to a chi-square distribution with $d = Tr(\mathbf{H}\boldsymbol{\Omega})^2/Tr\left[(\mathbf{H}\boldsymbol{\Omega})^2\right]$ degrees of freedom. Although these are not the asymptotic distributions of T_s and T_a, it has been repeatedly shown in simulation studies (e.g., Muthén, 1993; Satorra & Bentler, 1994) that these reference distributions closely approximate the actual (and presently unknown) distribution of these statistics.

Formal proofs of these results are given in Appendix B. It should pointed out that these are asymptotic results for the MNO model and therefore are applicable to the NOHARM model in as much as this is an approximation to MNO model. If one alternatively takes NOHARM's model as a model in itself, then appropriate substitutions on the above formulae should be made. The necessary adjustments to the derivatives can then be found in Christoffersson (1975, Appendix 1). However, we have found negligible differences in the estimated standard errors and goodness-of-fit tests when these adjusted derivative matrices are employed.

Large Sample Results for Rotated Factor Models

Following Browne and du Toit (1992), an unrestricted (a.k.a. exploratory) factor solution subject to the restrictions of a simplicity function, say Varimax, can be estimated directly from the second-order marginal proportions using instead of (14)

$$F_2^*(\boldsymbol{\theta}) = Min_{\theta}\left[\begin{pmatrix}\mathbf{p}_2\\ \mathbf{0}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})\\ \mathbf{C}(\boldsymbol{\theta})\end{pmatrix}\right]'\left[\begin{pmatrix}\mathbf{p}_2\\ \mathbf{0}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})\\ \mathbf{C}(\boldsymbol{\theta})\end{pmatrix}\right], \quad (32)$$

where $\mathbf{C}(\boldsymbol{\theta}) = \mathbf{0}$ denotes the set of restrictions imposed by the simplicity function on the model parameters. For instance, in the case of the Varimax criterion with Kaiser row normalization, $\mathbf{C}(\boldsymbol{\theta}) = \mathbf{0}$ consists (Archer & Jennrich, 1973; Browne & du Toit, 1992) of the following $p(p-1)/2$ equality constraints $\mathbf{z}_{ij} - \mathbf{z}_{ij} = \mathbf{0}$, $i < j, j = 2, \ldots, p$, where

$$\mathbf{Z} = \frac{1}{n}\boldsymbol{\Lambda}'\mathbf{D}_h^{-1}\boldsymbol{\Lambda}Diag(\boldsymbol{\Lambda}'\mathbf{D}_h^{-1}\boldsymbol{\Lambda}) - \boldsymbol{\Lambda}'\mathbf{D}_h^{-2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{(3)}. \quad (33)$$

In (33), $\boldsymbol{\Lambda}^{(3)}$ is obtained by cubing the elements of $\boldsymbol{\Lambda}$, and $\mathbf{D}_h$ is a diagonal matrix of communalities, i.e., $\mathbf{D}_h = Diag(\boldsymbol{\Lambda}\boldsymbol{\Lambda}')$.

Then, large sample standard errors and goodness-of-fit tests for unrestricted models subject to a single rotation criterion can be obtained via (27), (29), and (31) using

$$\boldsymbol{\Omega}^* = \begin{pmatrix}\boldsymbol{\Omega} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} \qquad \boldsymbol{\Delta}_{22}^* = \begin{pmatrix}\boldsymbol{\Delta}_{22}\\ \dfrac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\end{pmatrix} \quad (34)$$

in place of $\boldsymbol{\Omega}$ and $\boldsymbol{\Delta}_{22}$. Note that because in (32) $\mathbf{C}(\boldsymbol{\theta}) = \mathbf{0}$ act as identification constraints, it will have the same minimizer, subject to the constraints, as (14) expressed as a function of a lower echelon matrix $\boldsymbol{\Lambda}_L$.

However, the use of (32) to obtain rotated solutions is not recommended in practice except for small models, as the constraints implied by rotation simplicity function are so complex that minimizing (32) may be rather time consuming in large models. It is computationally more efficient to estimate rotated factor loadings as in the NOHARM program, where a low echelon matrix of factor loadings is estimated using (14) and these loadings are then rotated to a Varimax solution using a planar approach. In this case, we can also obtain standard errors and goodness-of-fit tests via (27), (29), and (31) using (34) and $\hat{F}_2$ given by (14).

These procedures may be used to obtain large sample results for rotated solutions that satisfy both orthogonal and oblique criteria (see Browne & du Toit, 1992). They can not be used to obtain standard errors for Promax rotated factor loadings as these do not satisfy a single simplicity function, but are instead the result of a two-stage rotation procedure.

A Numerical Example: The LSAT 7 Data

These data, originally analyzed by Bock and Lieberman (1970), have been analyzed using a variety of limited and full information methods among others by Christoffersson (1975), Muthén (1978), and Bock and Aitkin (1981) (for a review, see McDonald & Mok, 1985). For these data, McDonald (1997, pp. 266–268) reports using NOHARM for a one-factor solution, a two-factor solution after a Promax rotation, and a restricted two-factor solution. We have implemented the above formulae in Mathematica (Wolfram, 1999) where we used the Broyden-Fletcher-Godfarb-Shanno approach (BFGS; see Luenberger, 1984) to minimize (14). In Table 1 we report the one-factor and restricted two-factor solutions and provide standard errors for the parameter estimates and limited information goodness-of-fit tests of these models. We also provide a two-factor solution after Varimax rotation with Kaiser's row normalization with asymptotically correct standard errors for the rotated loadings. Our estimates are identical to those reported in McDonald (1997), except for the thresholds, which in McDonald (1997) appear with their signs reversed due to a typographical error.

If we compare our standard errors for the one-factor model with those obtained by Christoffersson (1975) and Muthén (1978), we observe that our standard errors for the factor loadings are somewhat larger than theirs. This is because their WLS methods are asymptotically efficient within the class of esti-

TABLE 1
NOHARM parameter estimates, estimated standard errors, and goodness-of-fit tests for the LSAT7 data

		One-Dimensional Solution	Two-Dimensional Varimax Solution		Two-Dimensional Restricted Solution	
Item	τ	λ	λ_1	λ_2	λ_1	λ_2
1	−0.946	0.492	0.814	0.148	0.819	0
	(0.047)	(0.064)	(0.394)	(0.054)	(0.372)	*(fixed)*
2	−0.407	0.536	0.185	0.496	0	0.555
	(0.041)	(0.061)	(0.139)	(0.150)	*(fixed)*	(0.067)
3	−0.745	0.718	0.215	0.796	0	0.784
	(0.044)	(0.067)	(0.133)	(0.189)	*(fixed)*	(0.086)
4	−0.269	0.418	0.317	0.272	0.248	0.241
	(0.040)	(0.056)	(0.181)	(0.134)	(0.184)	(0.160)
5	−1.007	0.370	0.310	0.222	0.267	0.178
	(0.048)	(0.067)	(0.181)	(0.138)	(0.207)	(0.182)

Model	T_s	df	*p*-value	T_a	df	*p*-value
One-dimensional	11.25	5	0.05	10.98	4.88	0.05
Two-dimensional unrestricted	0.64	1	0.42	0.64	1.00	0.42
Two-dimensional restricted	0.75	2	0.69	0.74	1.97	0.68

Notes: Asymptotic standard errors are in parentheses. The threshold estimates are the same for all models. In the two-dimensional restricted model, $\hat{\psi}_{21} = 0.471$ (0.235).

TABLE 2
Raw and adjusted residuals for the one-dimensional model for LSAT7 data

Item	1	2	3	4	5
1	—	−1.065	−1.997	2.132	1.822
2	−0.004	—	2.779	−0.870	−1.500
3	−0.006	0.006	—	−1.059	−0.001
4	0.010	−0.003	−0.003	—	0.127
5	0.008	−0.006	0	0.001	—

Notes: Raw residuals, $\hat{\epsilon}_{ij}$, below the diagonal; adjusted residuals, $\hat{a}_{ij}$, above the diagonal.

mators using first- and second-order information, whereas the ULS estimator implemented in NOHARM is not. However, the theoretical advantage of WLS estimators over ULS estimators is only realized when the sample size relative to model size is very large. Muthén (1993) has compared, in a simulation study, the estimated standard errors and goodness-of-fit tests of three-stage ULS versus WLS estimators in fitting a unidimensional normal ogive model. He has shown that for a one-factor model fitted to 15 skewed dichotomous variables, his ULS estimator yielded considerably less biased estimates and more accurate goodness-of-fit tests than his 1978 WLS estimator, even when the sample size was as large as 4,000. Furthermore, when the sample size was 1,000, the standard errors and goodness-of-fit tests of the WLS estimator were substantially off, whereas the ULS estimator gave reasonable results.

For completeness, in Table 2 we provide the raw and adjusted residuals, $\hat{\epsilon}_{ij}$ and $\hat{a}_{ij}$, after fitting the one-factor model to these data. The adjusted residuals indicate that the one-factor model does not adequately reproduce the bivariate proportions p_{23} and p_{14}. Also, for the one-factor model, it is straightforward to compute the pattern probabilities using (8) and compute full information X^2 and G^2 statistics. These are $X^2 = 32.26$ and $G^2 = 31.77$, both on 21 df, yielding $p = 0.056$ and $p = 0.062$, respectively.

Simulation Study

Consider the following two-factor model with eight indicators

$$\boldsymbol{\Lambda}' = \begin{pmatrix} 0.8 & 0.7 & 0.6 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.8 & 0.7 & 0.6 & 0.5 \end{pmatrix} \qquad \boldsymbol{\Psi} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

$$\boldsymbol{\tau}' = (0.5 \;\; 0.25 \;\; -0.25 \;\; -0.5 \;\; -0.5 \;\; -0.25 \;\; 0.25 \;\; 0.5).$$

Using this model, we perform a small simulation study to investigate the small sample behavior of NOHARM's estimation approach and our formula for its standard errors and goodness-of-fit tests. We compare the performance of this approach to that of

1) Two-stage approach with the bivariate normal integrals evaluated as in Divgi (1979), henceforth referred to as two-stage approach, rather than by (12) with $k = 3$ as in the two-stage approach implemented in NOHARM.
2) Muthén's (1993) three-stage ULS estimation procedure.

Two sample sizes were considered $N = 1{,}000$ and $N = 100$. One thousand replicates were employed under each condition, and to increase the comparability of the results, the same starting seed was used in all conditions. All necessary computations were performed using Mathematica, and the same minimization procedure, BFGS, was used in all conditions.

Standard errors and goodness-of-fit tests for Muthén's (1993) three-stage ULS estimation procedure can be obtained as follows: Letting $\boldsymbol{\kappa} = (\boldsymbol{\tau}, \boldsymbol{\rho})'$, Muthén (1978) showed that $\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}) \stackrel{a}{=} \tilde{\boldsymbol{\Delta}}^{-1}\sqrt{N}[\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\tau}, \boldsymbol{\theta})]$, where

$$\tilde{\boldsymbol{\Delta}} = \frac{\partial \boldsymbol{\pi}}{\partial \hat{\boldsymbol{\kappa}}'} = \begin{pmatrix} \boldsymbol{\Delta}_{11} & \mathbf{0} \\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{\rho} \end{pmatrix}.$$

Since

$$\tilde{\boldsymbol{\Delta}}^{-1} = \begin{pmatrix} \boldsymbol{\Delta}_{11}^{-1} & \mathbf{0} \\ -\boldsymbol{\Delta}_{\rho}^{-1}\boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{\rho}^{-1} \end{pmatrix},$$

it follows that the asymptotic distribution of the sample statistics used in Muthén's (1993) third stage (17) is

$$\sqrt{N}[\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\theta})] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Xi})$$

$$\boldsymbol{\Xi} = (-\boldsymbol{\Delta}_{\rho}^{-1}\boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}|\boldsymbol{\Delta}_{\rho}^{-1})\boldsymbol{\Gamma}(-\boldsymbol{\Delta}_{\rho}^{-1}\boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}|\boldsymbol{\Delta}_{\rho}^{-1})'. \tag{35}$$

Furthermore, when no restrictions are imposed on the thresholds, standard errors for its estimates can be obtained using (27), and standard errors for the estimated factor loadings and interfactor correlations can be obtained using

$$A\hat{c}ov(\hat{\boldsymbol{\theta}}) = \frac{1}{N}\left(\hat{\boldsymbol{\Delta}}_{\theta}'\hat{\boldsymbol{\Delta}}_{\theta}\right)^{-1}\hat{\boldsymbol{\Delta}}_{\theta}'\hat{\boldsymbol{\Xi}}\hat{\boldsymbol{\Delta}}_{\theta}\left(\hat{\boldsymbol{\Delta}}_{\theta}'\hat{\boldsymbol{\Delta}}_{\theta}\right)^{-1}. \tag{36}$$

Muthén (1993) has proposed statistics to assess the goodness of fit of the restrictions imposed by the model on the tetrachoric correlations. However, as he pointed out, the resulting tests are only meaningful if the distributional restrictions imposed by the model (i.e., dichotomized multivariate normality) hold. Unfortunately, the distributional assumptions underlying the use of tetrachoric correlations can only currently be assessed piecewise (for triplets of items) using the statistics proposed by Muthén and Hofacker (1988). Recently, Maydeu-Olivares (2001) has proposed a goodness-of-fit test that overcomes this shortcoming of the three-stage approach. Let $\hat{\boldsymbol{\epsilon}} := [\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}})]$. Maydeu-Olivares has shown that when Muthén's (1993) estimation procedure is used,

$$\sqrt{N}\hat{\boldsymbol{\epsilon}} = \sqrt{N}\left[\mathbf{p} - \boldsymbol{\pi}\left(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}}\right)\right] \xrightarrow{d} N\left(\mathbf{0}, \tilde{\mathbf{M}}\right) \qquad \tilde{\mathbf{M}} = \tilde{\mathbf{H}}\boldsymbol{\Gamma}\tilde{\mathbf{H}}' \tag{37}$$

where $\tilde{\mathbf{H}} = \mathbf{I}_{n(n+1)} - \tilde{\boldsymbol{\Delta}}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\tilde{\boldsymbol{\Delta}}^{-1}$ and, when no restrictions are imposed on τ,

$$\boldsymbol{\Delta} = \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_\theta \end{pmatrix}.$$

To assess the goodness of fit of the model to the first- and second-order marginals of the contingency table, Maydeu-Olivares (2001) proposed scaling $T = N\hat{\epsilon}'\hat{\epsilon}$ as follows

$$T_s = \frac{r}{Tr(\mathbf{M})} T \qquad T_a = \frac{Tr(\mathbf{M})}{Tr(\mathbf{M}^2)} T, \tag{38}$$

where T_s and T_a, denote the scaled (for mean) and adjusted (for mean and variance) test statistics. T_s is referred to a chi-square distribution with r degrees of freedom, whereas T_a is referred to a chi-square distribution with $d = Tr(\mathbf{M})^2/Tr(\mathbf{M}^2)$ degrees of freedom.

For the three estimation methods under consideration, we show in Table 3 the mean and standard deviation across replications of the estimated factor loadings and interfactor correlations along with the mean across replications of the estimated standard errors for these parameters. The results for the thresholds are not shown as the estimates and standard errors for these parameters are the same for the three methods under consideration. We observe in this table that the differences between the methods are rather small. Furthermore, no method appears to consistently outperform the others. Also, for this small model, we are able to obtain reasonable parameter estimates and standard errors with as few as 100 observations, although the standard errors obtained under this condition show a negative relative bias ranging from –4% to –11%. Of course, more accurate and less variable parameter estimates are obtained with the larger sample size. When $N = 1{,}000$, the standard errors no longer show a consistent negative bias.

In Table 4 we provide the empirical rejection rates of the proposed test statistics to assess the goodness of fit of the model at selected nominal rates. Their behavior is very similar across estimation methods. Again, for this small model the goodness-of-fit tests are remarkably accurate even when $N = 100$. Furthermore, the T_a statistic does not consistently outperform the simpler T_s statistic.

In sum, we have found that when the MNO model is fitted using the NOHARM approximating model using a two-stage approach, we obtain similar results as those obtained by fitting the MNO model itself using the same two-stage approach, or when fitting the MNO model using a three-stage approach. This is remarkable, as we have seen that the NOHARM model leads to a very poor approximation of the bivariate normal integrals. Given that all three approaches yield very similar results, it is interesting to compare them in terms of speed. In our implementation of these methods, the computation of the two-stage estimates when the NOHARM model was used was almost twice as fast as the computation

TABLE 3

Parameter estimates and estimated standard errors across 1,000 replications for an eight-variable, two-factor model (19 df)

		N = 1000									N = 100								
		NOHARM			Two-stage			Muthén's ULS			NOHARM			Two-stage			Muthén's ULS		
Par.	True	$\bar{x}$ est.	$\bar{x}$ SEs	s est.	$\bar{x}$ est.	$\bar{x}$ SEs	s est.	$\bar{x}$ est.	$\bar{x}$ SEs	s est.	$\bar{x}$ est.	$\bar{x}$ SEs	s est.	$\bar{x}$ est.	$\bar{x}$ SEs	s est.	$\bar{x}$ est.	$\bar{x}$ SEs	s est.
λ_{11}	0.8	0.802	0.043	0.043	0.800	0.043	0.043	0.801	0.043	0.044	0.798	0.135	0.144	0.797	0.135	0.144	0.803	0.136	0.148
λ_{21}	0.7	0.704	0.042	0.041	0.701	0.042	0.041	0.700	0.042	0.041	0.695	0.131	0.140	0.691	0.131	0.138	0.685	0.131	0.138
λ_{31}	0.6	0.600	0.044	0.044	0.601	0.044	0.044	0.601	0.044	0.045	0.597	0.137	0.147	0.597	0.137	0.147	0.595	0.137	0.147
λ_{41}	0.5	0.496	0.048	0.047	0.497	0.048	0.047	0.497	0.048	0.047	0.495	0.149	0.159	0.497	0.150	0.162	0.500	0.147	0.161
λ_{52}	0.8	0.803	0.043	0.044	0.801	0.043	0.044	0.801	0.043	0.044	0.802	0.132	0.139	0.800	0.132	0.138	0.807	0.134	0.142
λ_{62}	0.7	0.704	0.042	0.043	0.700	0.042	0.042	0.700	0.042	0.042	0.706	0.129	0.146	0.702	0.129	0.144	0.695	0.129	0.144
λ_{72}	0.6	0.599	0.044	0.045	0.600	0.044	0.045	0.600	0.044	0.045	0.596	0.135	0.143	0.597	0.136	0.143	0.595	0.135	0.142
λ_{82}	0.5	0.498	0.048	0.048	0.500	0.048	0.048	0.500	0.048	0.048	0.501	0.148	0.154	0.502	0.149	0.155	0.507	0.146	0.157
ψ_{21}	0.5	0.500	0.047	0.046	0.502	0.047	0.046	0.501	0.047	0.046	0.517	0.144	0.152	0.518	0.144	0.152	0.516	0.142	0.151

TABLE 4

Rejection rates at selected nominal alpha levels across 1,000 replications for an eight-variable, two-factor model (19 df)

	N = 1,000						N = 100					
	NOHARM		Two-stage		Muthén's		NOHARM		Two-stage		Muthén's	
α	T_s	T_a	T_s	T_a	T_s	T_a	T_s	T_a	T_s	T_a	T_s	T_a
1%	2.0	1.4	1.6	1.4	1.8	1.5	1.7	0.1	1.8	0.2	2.4	0.4
5%	6.3	4.9	5.8	4.8	6.6	5.2	6.3	3.7	6.2	3.9	7.2	4.8
10%	11.8	10.4	11.8	10.8	12.3	11.1	11.9	8.0	11.4	7.9	12.9	9.1
20%	21.9	21.1	21.4	20.6	22.4	21.2	20.4	17.3	20.2	16.9	22.4	19.6
30%	32.0	31.2	32.3	31.7	33.1	32.4	30.3	28.2	30.3	27.8	31.8	30.4
40%	41.8	41.8	41.4	41.0	42.5	42.5	39.5	39.5	39.4	39.2	41.9	41.8
50%	50.8	51.3	49.9	50.7	51.6	52.1	50.4	51.6	50.1	52.2	52.7	54.6
60%	60.7	62.3	60.6	61.5	62.0	62.7	59.9	63.4	60.1	63.1	62.7	65.5
70%	69.4	70.9	69.4	70.9	70.4	72.0	69.4	74.8	69.3	74.5	72.0	75.5
80%	79.3	80.7	79.0	80.1	79.2	80.6	80.2	84.9	80.1	84.7	81.8	85.6
90%	89.3	91.0	89.1	90.9	89.1	91.2	88.6	92.7	88.8	92.7	89.8	93.3

Notes: T_s and T_a are the scaled and adjusted test statistics.

of Muthén's (1993) estimates, and 17 times faster than when the two-stage estimates where obtained using Divgi's (1979) routine to evaluate the bivariate normal integrals.

Discussion and Conclusions

We have been able to provide asymptotic formulae for the standard errors of NOHARM's parameter estimates in both restricted models and unrestricted rotated models, as well as goodness-of-fit test of the model. The formulae presented here are adequate for small models with sample sizes as small as 100 observations. We have also shown that the results obtained using NOHARM are comparable to those obtained using Muthén's (1993) three-stage estimator. Furthermore, we found the estimation method implemented in NOHARM to be twice as fast as that of Muthén (1993), which is available, for instance, in LISREL 8 (Jöreskog & Sörbom, 1993).

Further work is needed to determine the minimum sample size needed to obtain reasonable results with models of more realistic size (e.g., 100 binary variables). Because NOHARM uses a two-stage estimation approach, it is possible to obtain standard errors and goodness-of-fit tests without storing any vector larger than $n(n-1)/2$ by recomputing quantities that one commonly would compute and store. These computational details will be given elsewhere. As a result, it is possible to obtain standard errors and goodness-of-fit tests for very large models. This is an added benefit over three-stage estimators which require storing the asymptotic covariance matrix of the second-stage estimates in order to obtain standard errors and goodness-of-fit tests.

In closing, we discuss a statistic that Gessaroli and de Champlain (1996) have proposed be used to assess the goodness of fit of normal ogive models estimated by NOHARM. This statistic is

$$GDCH = (N-3)\sum_{i=2}^{p}\sum_{j=1}^{i-1}\left(\tanh^{-1}\hat{r}_{ij}\right)^2, \tag{39}$$

where

$$\hat{r}_{ij} = \frac{\hat{\epsilon}_{ij}}{\sqrt{p_i(1-p_i)p_j(1-p_j)}}.$$

They further claim without proof an asymptotic chi-square distribution for (39). This statistic was originally proposed by Steiger (1980) for testing an identity matrix hypothesis for a correlation matrix. However, the $\hat{r}_{ij}$'s are not correlations, nor they are residual correlations as it is claimed by Gessaroli and de Champlain (1996) as the $\hat{\epsilon}_{ij}$'s are residual *uncentered* second-order joint moments of the data. Rather than investigating the limit behavior of (39), we have investigated its small sample behavior against the T_s and T_a statistics in the $N = 100$ condition of our simulation study. For the estimation procedure implemented in NOHARM, we found that the empirical rejection rates of (39) across 1,000 replications at nominal rejection rates $\alpha = \{1\%, 5\%, 10\%, 20\%\}$ were $\{0, 0.2, 0.4, 2.6\}$. Clearly, the behavior of (39) in small samples is unacceptable as it accepts the model almost invariably. It should be noted that this trend is already present in the simulation studies reported by Gessaroli and de Champlain (1996), where for the only nominal rejection rate they investigated ($\alpha = 5\%$), (39) accepts the null hypothesis all too often.

Appendix A: Derivation of π_i and π_{ij} from (8) and (10)

The expression for π_i is obtained as

$$\Pr(y_i = 1) = E[y_i] = E_\eta\{E[y_i|\eta]\}$$
$$= \Phi_1(-\tau_i) + \phi_1(\tau_i)\sum_{k=1}^{\infty}\frac{(\lambda_i'\Psi\lambda_i)^{k/2}}{k!}H_{k-1}(\tau_i)E_\eta\left[H_k\left(\frac{\lambda_i'\eta}{\lambda_i'\Psi\lambda_i}\right)\right]$$
$$= \Phi_1(-\tau_i).$$

This result follows from $E[y_i] = 1\Pr(y_i = 1) + 0\Pr(y_i = 0) = \Pr(y_i = 1)$, the double expectation theorem (e.g., Mittelhammer, 1996), $E_\eta\{E[y_i|\eta]\} = E[y_i]$, the expression for $E[y_i|\eta] = \Pr(y_i = 1|\eta)$ given by (10), and the properties of normalized Hermite polynomials (Stuart & Ord, 1987)

$$E[H_k(x)] = 0 \qquad E[H_k(x)H_{k'}(x)] = \begin{cases} k!, & k = k' \\ 0, & k \neq k' \end{cases} \tag{A1}$$

Similarly, (12) follows from

$$\Pr(y_i = 1, y_j = 1) = E[y_i, y_j] = E_\eta\{E[(y_i, y_j)|\boldsymbol{\eta}]\} = E_\eta\{E[y_i|\boldsymbol{\eta}]E[y_j|\boldsymbol{\eta}]\}$$

(10) and (A1), where $E[(y_i, y_j)\,|\,\boldsymbol{\eta}] = E[y_i\,|\,\boldsymbol{\eta}]E[y_j\,|\,\boldsymbol{\eta}]$ by the assumption of local independence.

Appendix B: Proofs of Key Results

To prove (28), we perform a Taylor series expansion of $\boldsymbol{\pi}_2(\boldsymbol{\theta}\,|\,\hat{\boldsymbol{\tau}})$ around $\boldsymbol{\tau} = \boldsymbol{\tau}_0$, obtaining $\boldsymbol{\pi}_2(\boldsymbol{\theta}\,|\,\hat{\boldsymbol{\tau}}) \overset{a}{=} \boldsymbol{\pi}_2(\boldsymbol{\tau}, \boldsymbol{\theta}) + \boldsymbol{\Delta}_{21}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})$. Now, from (19) and (26), $\sqrt{N}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \overset{d}{\to} \boldsymbol{\Delta}_{11}^{-1}\sqrt{N}\,[\mathbf{p}_1 - \boldsymbol{\pi}_1(\boldsymbol{\tau}_0)]$, and we may write

$$\sqrt{N}\begin{bmatrix}\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\tau},\boldsymbol{\theta}) \\ \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}}) - \boldsymbol{\pi}_2(\boldsymbol{\tau},\boldsymbol{\theta})\end{bmatrix} \overset{d}{\to} \begin{pmatrix}\mathbf{0} & \mathbf{I} \\ \boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}^{-1} & \mathbf{0}\end{pmatrix}\sqrt{N}\begin{bmatrix}\mathbf{p}_1 - \boldsymbol{\pi}_1(\boldsymbol{\tau}) \\ \mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\tau},\boldsymbol{\theta})\end{bmatrix}. \quad \text{(B1)}$$

Also,

$$\sqrt{N}[\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})] \overset{d}{\to} \sqrt{N}(\mathbf{I}|-\mathbf{I})\begin{bmatrix}\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\tau},\boldsymbol{\theta}) \\ \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}}) - \boldsymbol{\pi}_2(\boldsymbol{\tau},\boldsymbol{\theta})\end{bmatrix}$$
$$= (-\boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}^{-1}|\mathbf{I}_p)\sqrt{N}\begin{bmatrix}\mathbf{p}_1 - \boldsymbol{\pi}_1(\boldsymbol{\tau}) \\ \mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\tau},\boldsymbol{\theta})\end{bmatrix} \quad \text{(B2)}$$

Equation (28) follows from (B2) and (24).

Let $\boldsymbol{\epsilon} := \mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}\,|\,\hat{\boldsymbol{\tau}})$, to prove (29), from (14) we regard NOHARM's second-stage estimator $\hat{\boldsymbol{\theta}}$ as a solution to

$$\frac{\partial F(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})}{\partial\boldsymbol{\theta}'} = \frac{\partial\boldsymbol{\epsilon}}{\partial\boldsymbol{\theta}'}\frac{\partial F(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})}{\partial\boldsymbol{\epsilon}'} = -\boldsymbol{\Delta}_{22}^{*}(2\boldsymbol{\epsilon}) = \mathbf{0},$$

where $\boldsymbol{\Delta}_{22}^{*}$ equals $\boldsymbol{\Delta}_{22}$ evaluated at $\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}$, the first-stage parameter estimates. That is, $\hat{\boldsymbol{\theta}}$ must satisfy

$$\mathbf{p}_2 - \boldsymbol{\pi}_2(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\tau}}) = \mathbf{0}. \quad \text{(B3)}$$

Now, a linear Taylor expansion of $\boldsymbol{\pi}_2(\hat{\boldsymbol{\theta}}\,|\,\hat{\boldsymbol{\tau}})$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ yields

$$\boldsymbol{\pi}_2(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\tau}}) \overset{a}{=} \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}}) + \boldsymbol{\Delta}_{22}^{*}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \quad \text{(B4)}$$

Substituting this into (B3) and rearranging terms, $\boldsymbol{\Delta}_{22}^{*}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{a}{=} \mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}\,|\,\hat{\boldsymbol{\tau}})$, so that

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{a}{=} \left(\boldsymbol{\Delta}_{22}^{*\prime}\boldsymbol{\Delta}_{22}^{*}\right)^{-1}\boldsymbol{\Delta}_{22}^{*\prime}\sqrt{N}[\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})]. \quad \text{(B5)}$$

Combining (B5) with (28), we have

$$Acov(\hat{\boldsymbol{\theta}}) = \frac{1}{N}\left(\boldsymbol{\Delta}_{22}^{*\prime}\boldsymbol{\Delta}_{22}^{*}\right)^{-1}\boldsymbol{\Delta}_{22}^{*\prime}\boldsymbol{\Omega}\boldsymbol{\Delta}_{22}^{*}\left(\boldsymbol{\Delta}_{22}^{*\prime}\boldsymbol{\Delta}_{22}^{*}\right)^{-1}, \quad \text{(B6)}$$

which can be estimated by evaluating all derivative matrices at $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}})$ and substituting sample proportions for probabilities to consistently estimate $\boldsymbol{\Gamma}$, yielding (29).

Now, to prove (30), we note that

$$\hat{\boldsymbol{\epsilon}} := \mathbf{p}_2 - \boldsymbol{\pi}_2(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\tau}}) = (\mathbf{I}| - \mathbf{I})\begin{bmatrix}\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}}) \\ \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}}) - \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})\end{bmatrix}$$

and that from (B4)

$$\sqrt{N}\begin{bmatrix}\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}}) \\ \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}}) - \boldsymbol{\pi}_2(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\tau}})\end{bmatrix} \overset{d}{\to} \begin{bmatrix}\mathbf{I} \\ \boldsymbol{\Delta}_{22}^{*}\left(\boldsymbol{\Delta}_{22}^{*\prime}\boldsymbol{\Delta}_{22}^{*}\right)^{-1}\boldsymbol{\Delta}_{22}^{*\prime}\end{bmatrix}\sqrt{N}[\mathbf{p}_2 - \boldsymbol{\pi}_2(\boldsymbol{\theta}|\hat{\boldsymbol{\tau}})] \quad \text{(B7)}$$

Finally, using Theorem 2.1 of Box (1954) and (30),

$$T := N\hat{F}_2 = N\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} \overset{d}{\to} \sum_{i=1}^{j}\alpha_i\chi_1^2, \quad \text{(B8)}$$

where the χ_1^2's are independent chi-square variables with one degree of freedom and the α_i's are the non-null eigenvalues of $\left[\mathbf{I} - \boldsymbol{\Delta}_{22}^{*}\left(\boldsymbol{\Delta}_{22}^{*\prime}\boldsymbol{\Delta}_{22}^{*}\right)^{-1}\boldsymbol{\Delta}_{22}^{*\prime}\right]\boldsymbol{\Omega}$.

References

Archer, C. O. & Jennrich, R. I. (1973). Standard errors for rotated factor loadings. *Psychometrika, 38,* 581–592.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35,* 179–197.

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

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics, 16,* 769–771.

Browne, M. W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

Browne, M. W., & du Toit, S. H. C. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research, 27,* 269–300.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40,* 5–32.

David, F. N. (1953). A note on the evaluation of the multivariate normal integral. *Biometrika, 40*, 458–459.

Divgi, D. R. (1979). Calculation of univariate and bivariate normal probability functions. *Annals of Statistics, 7*, 903–910.

Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research, 23*, 267–269.

Gessaroli, M. E., & de Champlain, A. F. (1996). Using an approximate chi-square statistic to test the number of dimensions underlying the responses to a set of items. *Journal of Educational Measurement, 33*, 157–192.

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to simple structure. *British Journal of Statistical Psychology, 17*, 65–70.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8. User's reference guide.* Chicago, IL: Scientific Software.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187–200.

Kendall, M. G. (1941). Relations connected with the tetrachoric series and its generalization. *Biometrika, 32*, 196–198.

Knol, D. L., & Berger, M. P. F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26*, 457–477.

Luenberger, D. G. (1984). *Linear and non-linear programming.* Reading, MA: Addison-Wesley.

Maydeu-Olivares, A. (2001). Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgment sampling. *Psychometrika, 66*, 209–227.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometric Monographs, No. 15.*

McDonald, R. P. (1982a). Linear vs. nonlinear models in latent trait theory. *Applied Psychological Measurement, 6*, 379–396.

McDonald, R. P. (1982b). Fitting latent trait models. In D. Spearrit (Ed.), *Some alternative approaches for the improvement of measurement in education and psychology.* Australian Council for Educational Research.

McDonald, R. P. (1985). Unidimensional and multidimensional models for item response theory. In D. J. Weiss (Ed.), *Proceedings of the 1982 item response and computerized adaptive testing conference.* Minneapolis: University of Minnesota.

McDonald, R. P. (1997). Normal ogive multidimensional model. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 257–269). New York: Springer.

McDonald, R. P., & Mok, M. C. (1995). Goodness of fit in item response models. *Multivariate Behavioral Research, 54*, 483–495.

Mittelhammer, R. C. (1996). *Mathematical statistics for economics and business.* New York: Springer.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.

Muthén, B. (1993). Goodness of fit with categorical and other non normal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205–234). Newbury Park, CA: Sage.

Muthén, B., du Toit, S. H. C. & Spisic, D. (in press). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika.*

Muthén, B. & Hofacker, C. (1988). Testing the assumptions underlying tetrachoric correlations. *Psychometrika, 53*, 563–578.

Rao, C. R. (1973). *Linear statistical inference and its applications.* New York: Wiley.

Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271–286). New York: Springer.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*, 131–151.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye and C. C. Clogg (Eds.), *Latent variable analysis. Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics, Vol. 1* (5th ed). New York: Oxford.

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

ten Berge, J. M. F., Knol, D. L., & Kiers, H. A. L. (1988). A treatment of the orthomax rotation family in terms of diagonalization, and a re-examination of a singular value approach to Varimax rotation. *Computational Statistics Quarterly, 3*, 207–217.

Terza, J. V., & Welland, U. (1991). A comparison of bivariate normal algorithms. *Journal of Statistical Computation and Simulation, 39*, 115–127.

Teugels, J. L. (1990). Some representations of the multivariate Bernoulli and binomial distributions. *Journal of Multivariate Analysis, 32*, 256–268.

van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory.* New York: Springer.

Wolfram, S. (1999). *Mathematica 4.* Champaign, IL: Wolfram Research.

Author

ALBERT MAYDEU-OLIVARES is Associate Professor, Faculty of Psychology, University of Barcelona, P. Valle de Hebron, 171, 08035 Barcelona, Spain; amaydeu@tinet.fut.es.