PSYCHOMETRIKA—VOL. 66, NO. 2, 209–228
JUNE 2001

LIMITED INFORMATION ESTIMATION AND TESTING OF THURSTONIAN MODELS FOR PAIRED COMPARISON DATA UNDER MULTIPLE JUDGMENT SAMPLING

ALBERT MAYDEU-OLIVARES

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

We relate Thurstonian models for paired comparisons data to Thurstonian models for ranking data, which assign zero probabilities to all intransitive patterns. We also propose an intermediate model for paired comparisons data that assigns nonzero probabilities to all transitive patterns and to some but not all intransitive patterns.

There is a close correspondence between the multidimensional normal ogive model employed in educational testing and Thurstone's model for paired comparisons data under multiple judgment sampling with minimal identification restrictions. Alike the normal ogive model, Thurstonian models have two formulations, a factor analytic and an IRT formulation. We use the factor analytic formulation to estimate this model from the first and second order marginals of the contingency table using estimators proposed by Muthén. We also propose a statistic to assess the fit of these models to the first and second order marginals of the contingency table. This is important, as a model may reproduce well the estimated thresholds and tetrachoric correlations, yet fail to reproduce the marginals of the contingency table if the assumption of multivariate normality is incorrect.

A simulation study is performed to investigate the performance of three alternative limited information estimators which differ in the procedure used in their final stage: unweighted least squares (ULS), diagonally weighted least squares (DWLS), and full weighted least squares (WLS). Both the ULS and DWLS show a good performance with medium size problems and small samples, with a slight better performance of the ULS estimator.

Key words: UMD, EWMD, WMD, GLS estimation, LISREL, categorical data analysis, preference data, MPLUS, binary data, goodness of fit.

1. Introduction

Consider the problem of modeling the choice behavior of a homogeneous population of subjects in a paired comparisons design. For a fixed set of n objects and a random sample of N individuals from the population of interest, this experimental design consists in constructing all possible pairs of objects,

$$\tilde{n} = \binom{n}{2} = \frac{n(n-1)}{2},$$

and presenting them one pair at a time to each individual in the sample. These individuals are also given some preference or choice criterion and are asked to express their preferences for one object in each pair using the specified criterion.

Throughout this presentation, we shall assume that: (a) no equality judgments are allowed, (b) each subject in the sample is asked to judge all pairs—this is what Bock and Jones (1968) refer to as *multiple judgment sampling*—and that (c) presentation order effects are negligible—possibly through randomization of the order of presentation of pairs, and of the stimuli within a pair.

This paper is based on the author's doctoral dissertation; Ulf Böckenholt, advisor. The final stages of this research took place while the author was at the Department of Statistics and Econometrics, Universidad Carlos III de Madrid. The author is indebted to Adolfo Hernández for stimulating discussions that helped improve this paper, and to Ulf Böckenholt and the Associate Editor for a number of helpfulsuggestions to a previous draft.

Requests for reprints should be sent to Albert Maydeu-Olivares, Faculty of Psychology, University of Barcelona, Passeig Vall d'Hebron, 171, 08035 Barcelona SPAIN. E-Mail: amaydeu@psi.ub.es

0033-3123/2001-2/1998-0681-A $00.75/0

One of the most widely known class of models for paired comparisons data are Thurstonian models which assume that (Thurstone, 1927): (a) whenever a pair of stimuli is presented to a subject it elicits a continuous preference (utility function, or in Thurstone's terminology, *discriminal process*) for each stimulus; (b) the stimulus whose value is larger at the moment of the comparison will be preferred by the subject; (c) these unobserved preferences are normally distributed in the population.

By imposing constraints on the mean and covariance matrix of the underlying normal density one obtains different Thurstonian models such as Case III and Case V models (Thurstone, 1927), the factorial model (Takane, 1980; Heiser & de Leeuw, 1981), the wandering vector model (Carroll, 1980; De Soete & Carroll, 1983), the ideal point model (Brady, 1989), or the wandering ideal point model (De Soete, Carroll & DeSarbo, 1986). Takane (1987) discussed what restrictions these models impose on the binary pattern probabilities. He also proposed a mean and covariance structures analysis approach to estimating these models via Muthén's (1978) limited information estimator. However, Takane (1987) provided neither identification restrictions nor empirical examples and hence his paper largely remained programmatic. Here, taking on Takane's (1987) suggestive proposals, we provide a set of identification restrictions for Thurstonian paired comparisons models and we discuss the estimation of these models by the limited information methods proposed by Muthén (1978, 1993; Muthén, du Toit & Spisic, in press).

A drawback of these limited information methods is that no statistical test for assessing the overall fit of the models to the contingency table was available. A test statistic will be proposed to overcome this limitation. Also, a small simulation study will be performed to investigate the small sample behavior of this statistic and the relative merits in estimating these models of three different estimators proposed by Muthén (1978, 1993; Muthén, du Toit & Spisic, in press).

Thurstonian models for paired comparisons data assign nonzero probabilities to all paired comparisons patterns. These models are closely related to Thurstonian models for ranking data which can be seen as models for paired comparisons data that assign nonzero probability only to transitive paired comparisons (Maydeu-Olivares, 1999). A model between these two extremes that may be of interest in applications would postulate that intransitivities should occur only for certain combinations of objects. We refer to these models as models with *localized intransitivities* and discuss how they can be specified within a Thurstonian framework.

Most of our discussion revolves around the unrestricted Thurstonian (UT) model for paired comparisons data. This is a Thurstonian model in which the mean and covariance matrix of the underlying normal variates are unconstrained except for minimal restrictions needed for its identification. We find this model appealing in applications because it has a straightforward interpretation. Also, it is a useful benchmark model to assess the suitability of the whole class of Thurstonian models to the data at hand. With the identification restrictions provided here, there are close similarities between the UT model and the multidimensional normal ogive (MNO) model employed in the educational testing literature. Just as the MNO model can be expressed using a factor analytic (FA) or an item response theory (IRT) formulation (Takane & de Leeuw, 1987), so does the UT model, and we shall provide both formulations of the model.

2. Thurstone's Model for Multiple Judgment Paired Comparison Data

Consider a set of n objects and a random sample of N individuals; $\tilde{n}$ pairs of objects are constructed and each pair is presented to each individual in the sample. The outcomes of each paired comparison will be represented by a dichotomous random variable y_l indicating whether for each ordered pairwise combination of objects l a subject chooses object i or i'

$$y_{lj} = \begin{cases} = 1 & \text{if subject } j \text{ chooses object } i \\ = 0 & \text{if subject } j \text{ chooses object } i' \end{cases} \qquad l = 1, \ldots, \tilde{n};\ j = 1, \ldots, N \tag{1}$$

where $l \equiv (i, i')$, $(i = 1, \ldots, n-1; i' = i+1, \ldots, n)$.

Let t_{ij} denote subject j's unobserved continuous preference for object i. Thurstone (1927) proposed performing the following pairwise linear transformation on the set of unobserved preferences

$$y_l^* = t_i - t_{i'} \tag{2}$$

Since under the model choices are performed according to the rule:

$$\text{Choose object } \begin{cases} i & \text{if} \quad t_{ij} \geq t_{i'j} \\ i' & \text{if} \quad t_{ij} < t_{i'j} \end{cases},$$

it follows that

$$y_{lj} = \begin{cases} = 1 & \text{if} \quad y_{lj}^* \geq 0 \\ = 0 & \text{if} \quad y_{lj}^* < 0 \end{cases}. \tag{3}$$

Now, as proposed by Takane (1987), Thurstonian models for $\tilde{n}$-dimensional patterns of paired comparisons are obtained by considering the joint distribution of a n-dimensional vector of unobserved continuous preferences $\mathbf{t}$ and a $\tilde{n}$-dimensional vector of random errors $\mathbf{e}$ associated with each specific paired comparison. We assume that

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{e} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_t \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_t & \\ \mathbf{0} & \boldsymbol{\Omega} \end{pmatrix} \right) \tag{4}$$

where $\boldsymbol{\Omega}$ is a diagonal matrix with diagonal elements ω_l.

As in (2) we assume the linear transformation

$$\mathbf{y}^* = \begin{pmatrix} \mathbf{A} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{t} \\ \mathbf{e} \end{pmatrix} = \mathbf{A}\,\mathbf{t} + \mathbf{e}, \tag{5}$$

where $\mathbf{A}$ is a $\tilde{n} \times n$ matrix of simple pairwise contrasts with element in row l and column k

$$a_{lk} = \begin{cases} 0 & \text{if} \quad k \notin \{i, i'\} \\ 1 & \text{if} \quad k = i \\ -1 & \text{if} \quad k = i' \end{cases}.$$

For example, when $n = 4$, $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}. \tag{6}$$

Finally from (3), for any paired comparisons pattern

$$\Pr\left[\bigcap_{l=1}^{\tilde{n}} y_l \right] = \int \cdots \int_{\mathbf{R}} \phi_{\tilde{n}}(\mathbf{y}^* : \mathbf{A}\boldsymbol{\mu}_t, \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}' + \Omega) d\mathbf{y}^*, \tag{7}$$

where $\phi_{\tilde{n}}(\bullet)$ denotes a $\tilde{n}$-dimensional normal density, and the limits of integration in (7) are

$$R_l = \begin{cases} (0, \infty) & \text{if} \quad y_l = 1 \\ (-\infty, 0) & \text{if} \quad y_l = 0 \end{cases}. \tag{8}$$

Equations (7) and (8) define an unrestricted Thurstonian model. Restricted Thurstonian models can be obtained by imposing suitable constraints on $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ (see Takane, 1987).

3. Identification Restrictions

Since the observed variables are dichotomous, the means and variances of $\mathbf{y}^*$ are not identifiable separately. This can be solved for instance by setting $\sigma^2_{y_l^*} = 1, \forall l$, so that $\mathbf{y}^*$ has a correlation structure, $\mathbf{P}_{y^*}$. One way to enforce the variance normalization restriction on $\mathbf{y}^*$ is by using a restriction analogous to that employed in factor analysis for dichotomous variables

$$\boldsymbol{\Omega} = \mathbf{I} - \text{diag}(\mathbf{A}\Sigma_t\mathbf{A}'). \tag{9}$$

Hence, Thurstonian models will be identified if and only if they are can be identified from the reduced form parameters of the model, $\boldsymbol{\mu}_{y^*}$ and $\mathbf{P}_{y^*}$. We notice that since $\mathbf{A}$ is of rank $n-1$, the parameters of any Thurstonian model suffer from a location indeterminacy. In the case of the UT model, this indeterminacy can be solved for instance by setting

$$\begin{aligned} \mu_n &= 0 \\ \sigma_{ii} &= 1, \forall i \end{aligned} \tag{10}$$

so that $\mathbf{P}_t$, a correlation matrix among the continuous unobserved preferences, is estimated instead of their covariance matrix.

Consider now Thurstone's (1927) Case V and Case III models, in which $\boldsymbol{\Sigma}_t = \psi\mathbf{I}$ and $\boldsymbol{\Sigma}_t$ is a diagonal matrix, respectively. Given (9), the location indeterminacy in these models can be solved (and hence these models are identified) by just letting $\mu_n = 0$.

4. Relationships Between the Unrestricted Thurstonian Model for Paired Comparisons Data and the Normal Ogive Model for Item Response Modeling

The UT model for paired comparisons data is a latent trait model because the random variables $\mathbf{y}^*$ conditional on $\mathbf{t}$ are mutually independent by the diagonal assumption on $\boldsymbol{\Omega}$. Furthermore, if we define item response models as latent trait models with categorical indicators, then this model is an item response model. We shall now see that there is a close relationship between the UT model defined by (7) and (8) with identification restrictions (9) and (10) and the multidimensional normal ogive (MNO) item response model for dichotomous test data (e.g., Christoffersson, 1975). The latter is defined by

$$\Pr\left[\bigcap_{i=1}^{n} y_i\right] = \int_{\widehat{\mathbf{R}}} \phi_n(\mathbf{w}^*; \boldsymbol{\mu}_{w^*}, \mathbf{P}_{w^*})\, d\mathbf{w}^* \tag{11}$$

$$\widehat{R}_i = \begin{cases} (\tau_i, \infty) & \text{if} \quad y_i = 1 \\ (-\infty, \tau_i) & \text{if} \quad y_i = 0 \end{cases} \qquad i = 1, \ldots, n.$$

In (11), it is assumed that $\mathbf{w}*$ has the following mean and correlation structure: $\boldsymbol{\mu}_{w^*} = \mathbf{0}$ and $\mathbf{P}_{w^*} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$, and it is further assumed that $\mathbf{w}*$ has been dichotomized according to a set of thresholds $\boldsymbol{\tau}$.

There is a rather obvious relationship between the reduced form parameters of this model, $\boldsymbol{\tau}$ and $\mathbf{P}_{w^*}$, with those of the UT model for paired comparisons data.

Lemma. The reduced forms of the MNO model for $\tilde{n}$ items and of the UT model for n objects are equivalent with

$$\begin{aligned} \boldsymbol{\mu}_{y^*} &= -\boldsymbol{\tau} \\ \mathbf{P}_{y^*} &= \mathbf{P}_{w^*} \end{aligned} \tag{12}$$

Proof. Perform a change of variable $y_l^* = w_l^* - \tau_l$, in the normal ogive model. Then $dy_l^* = dw_l^*$ and at $w_l^* = \tau_l$, $y_l^* = 0$. □

As for the relationships between the actual parameters of both models, the MNO model can be identified (McDonald, 1985) by letting $\mathbf{\Lambda}$ be a low echelon matrix, $\mathbf{\Phi}$ a correlation matrix, and $\mathbf{\Psi} = \mathbf{I} - \text{diag}(\mathbf{\Lambda\Phi\Lambda}')$. Taking into account these identification restrictions, we see that the MNO and UT models differ in

1. The dimension of the latent variable vector. This is fixed in the UT model, n (in fact, $n-1$, see the next section), whereas the MNO model just requires that $p < n$.
2. The correlation structure. $\mathbf{\Lambda}$ in the MNO model consists of regression coefficients to be estimated, whereas $\mathbf{A}$ in the UT model is a fixed design matrix.
3. The mean structure. This is an unrestricted vector, $-\boldsymbol{\tau}$, in the MNO model, whereas in the UT model it is a linear function (given by the design matrix $\mathbf{A}$) of the parameter vector $\boldsymbol{\mu}_t$.

Equation (11) corresponds to a factor analytic formulation of the MNO model (Takane & de Leeuw, 1987). This model has an alternative formulation that is more frequently used in item response theory (IRT). Just as the MNO model has two formulations, so does the unrestricted Thurstonian model. We next present an IRT formulation of the UT model.

5. An IRT Formulation of the Unrestricted Thurstonian Model for Paired Comparison Data

The identification restrictions (10) are not unique of course and were chosen for ease of interpretation (see Maydeu-Olivares, 1999, Appendix B). Now, to formulate the UT model as an IRT model it is convenient to reparameterize it as

$$\mathbf{z} = \mathbf{S}\,\mathbf{t} \tag{13}$$

where $\mathbf{S} = [\mathbf{I}_{n-1} | -\mathbf{1}_{n-1}]$ is a matrix arbitrarily chosen so that its rows are in the row space of $\mathbf{A}$. With this reparameterization, the parameters being estimated are $\boldsymbol{\mu}_z = \mathbf{S}\boldsymbol{\mu}_t$ and $\mathbf{\Sigma}_z = \mathbf{S}\mathbf{P}_t\mathbf{S}'$. Letting $\mu_i \in \boldsymbol{\mu}_t$, $\tilde{\mu}_i \in \boldsymbol{\mu}_z$, $\rho_{ii'} \in \mathbf{P}_t$, $\tilde{\sigma}_{ii'} \in \mathbf{\Sigma}_z$, the relationship between both parameterizations is given by

$$\begin{aligned}
\tilde{\mu}_i &= \mu_i \\
\tilde{\sigma}_{ii} &= 2 - 2\rho_{ni} \qquad i = 1, \ldots, n-1; i' = i+1, \ldots, n \\
\tilde{\sigma}_{ii'} &= 1 + \rho_{ii'} - \rho_{ni} - \rho_{ni'}.
\end{aligned} \tag{14}$$

With this reparameterization, (7) can be rewritten as

$$\Pr\left[\bigcap_{l=1}^{\tilde{n}} y_l\right] = \int_{\mathbf{R}} \cdots \int \phi_{\tilde{n}}(\mathbf{y}^* : \mathbf{K}\boldsymbol{\mu}_z, \mathbf{K}\mathbf{\Sigma}_z\mathbf{K}' + \mathbf{\Omega})\, d\mathbf{y}^* \tag{15}$$

where $\mathbf{K} = \mathbf{A}\mathbf{S}'(\mathbf{S}\mathbf{S}')^{-1}$ equals the first $n-1$ columns of $\mathbf{A}$, and now $\mathbf{\Omega} = \mathbf{I} - \text{diag}(\mathbf{K}\mathbf{\Sigma}_z\mathbf{K}')$. Now, as in Takane and de Leeuw (1987) we let

$$\begin{aligned}
\Pr\left[\bigcap_{l=1}^{\tilde{n}} y_l\right] &= \int_{\mathbf{R}} \cdots \int \left\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_{\tilde{n}}(\mathbf{y}^*|\mathbf{z})\phi_{n-1}(\mathbf{z})\, d\mathbf{z} \right\} d\mathbf{y}^* \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_{n-1}(\mathbf{z} : \mathbf{0}, \mathbf{\Sigma}_z) \left\{ \int_{\mathbf{R}} \cdots \int \phi_{\tilde{n}}(\mathbf{y}^*|\mathbf{z} : \boldsymbol{\mu}_{y^*} - \mathbf{K}\mathbf{z}, \mathbf{\Omega})\, d\mathbf{y}^* \right\} d\mathbf{z}
\end{aligned}$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_{n-1}(\mathbf{z} : \mathbf{0}, \boldsymbol{\Sigma}_z) \prod_{l}^{\tilde{n}} \left(\int_{R_l} \phi_1(y_l^* | \mathbf{z} : \mathbf{k}_l'\mathbf{z} + \mu_{y_l^*}, \omega_l)\, dy_l^* \right) d\mathbf{z}. \quad (16)$$

In the last expression in (16) we use that the $y_l^*|\mathbf{z}$ are mutually uncorrelated by the diagonality assumption on $\boldsymbol{\Omega}$, and hence independent by the multivariate normality of $\mathbf{y}|\mathbf{z}$. Finally,

$$\int_{R_l} \phi_1\left(y_l^* | \mathbf{z} : \mathbf{k}_l'\mathbf{z} + \mu_{y_l^*}, \omega_l\right) dy_l^* = \left(\Phi_1 \left(\frac{\mathbf{k}_l'\mathbf{z} + \mu_{y_l^*}}{\sqrt{\omega_l}} \right) \right)^{y_l} \left(1 - \Phi_1 \left(\frac{\mathbf{k}_l'\mathbf{z} + \mu_{y_l^*}}{\sqrt{\omega_l}} \right) \right)^{1-y_l} \quad (17)$$

where $\Phi_1(\bullet)$ denotes a univariate standard normal distribution function.

Equation (16) with (17) is the IRT formulation of the unrestricted Thurstonian model for paired comparisons data in which pattern probabilities require integrating over the $n-1$ dimensional multivariate normal density $\mathbf{z}$. The purpose of the reparameterization (13) is to reduce the dimensionality of the integration problem from n to $n-1$.

6. Model Interpretation, Improper Solutions, Boundary Solutions, and Models with Localized Intransitivities

Under the sampling scheme considered in this paper, the parameters of the unrestricted Thurstonian model have a straightforward interpretation: The μ_i and σ_{ii} denote the mean and variance of the population's unobserved continuous preferences for that object, while $\rho_{ii'}$ denotes the correlation between the preferences for objects i and i' in that population.

The random errors $\mathbf{e}$ in (5) are crucial in modeling paired comparisons data. Their inclusion allows the modeling of intransitive patterns of paired comparisons. A pattern of binary preferences is said to be *transitive* when given the pattern it is possible to rank order the objects, and intransitive otherwise. Substantively, a random error $e_{(i,i')j}$ reflects that a subject's preference for an object can change during the paired comparisons experiment as the object is presented next to different objects, thus giving rise to intransitivities. These errors are assumed to be uncorrelated with the continuous preferences and uncorrelated with each other, so that their covariance matrix, $\boldsymbol{\Omega}$, is diagonal.

Maydeu-Olivares (1999) has shown that since $\mathbf{A}$ is of rank $n-1$, $\mathbf{AP}_t\mathbf{A}'$ has rank $n-1$ and that (7) and (8) with $\boldsymbol{\Omega} = \mathbf{0}$ assign nonzero probabilities only to transitive patterns, thus defining a Thurstonian model suitable for ranking data. Mathematically, the addition of $\boldsymbol{\Omega}$ to $\mathbf{AP}_t\mathbf{A}'$ transforms $\mathbf{P}_{y^*}$ into a positive definite matrix, so that the model assigns nonzero probabilities to all $2^{\tilde{n}}$ paired comparisons patterns.

Therefore, if all members of a population were transitive in their paired comparisons preferences, then $\boldsymbol{\Omega} = \mathbf{0}$. We have found in unpublished applications that when all objects to be compared are similar to each other (as when the objects to be compared are different products within a market segment), almost invariably over 85% of the subjects in the sample yield transitive patterns. In these situations, one should expect some diagonal elements of $\boldsymbol{\Omega}$ be very close to zero in the population. In fact, any number of them may be zero in the population as long as

$$\mathbf{P}_{y^*} = \mathbf{AP}_t\mathbf{A}' + \boldsymbol{\Omega}$$

is positive definite, since that is a necessary and sufficient condition for the model to assign nonzero probabilities to all paired comparisons patterns.

Because some diagonal elements of $\boldsymbol{\Omega}$ can be zero in the population, improper solutions in which some estimates become negative are very likely to occur in paired comparisons applications with few observed intransitive patterns. Now, since with the identification restrictions laid out in section 3, $\boldsymbol{\Omega} = \mathbf{I} - \text{diag}(\mathbf{AP}_t\mathbf{A}')$, its elements are of the form $\omega_l = 1 - 2\rho_{ii'}$. Therefore,

for ω_l to be positive,

$$|\rho_{ii'}| > \frac{1}{2}, \quad \forall \rho_{ii'} \in \mathbf{P}_t. \tag{18}$$

Hence, the $\rho_{ii'}$ are not properly correlations since they are bounded by (18) and they should be interpreted accordingly.

Should an improper solution arise, inequality constraints can be enforced by reparameterization (McDonald, 1980) to obtain a boundary solution. However, whenever a boundary solution is obtained, the rank of $\mathbf{P}_{y^*}$ should be evaluated to determine whether the model assigns nonzero probabilities to all paired comparisons patterns. The rank of $\mathbf{P}_{y^*}$ (i.e., its maximum rank over the admissible parameter space) can be efficiently determined using computer algebra (see Bekker, Merckens & Wansbeek, 1994). Direct verification shows, for example, that for $\mathbf{P}_{y^*}$ to be positive definite when $n = 3$, only one element of $\boldsymbol{\omega}$ needs to be nonzero. On the other hand, when $n = 4$, (a) if 4 or more elements of $\boldsymbol{\omega}$ are nonzero, $\mathbf{P}_{y^*}$ will be positive definite; (b) if 2 or fewer elements are nonzero, $\mathbf{P}_{y^*}$ will not be positive definite, and (c) if there are exactly three nonzero elements, $\mathbf{P}_{y^*}$ will be positive for all but four permutations of the zero elements in $\boldsymbol{\omega}$. Note that a necessary condition for $\mathbf{P}_{y^*}$ to be of full rank is that at least $\tilde{n} - n + 1$ elements of $\boldsymbol{\omega}$ are nonzero.

Whenever the UT model has zero elements in $\boldsymbol{\omega}$, $\mathbf{y}^*|\mathbf{z}$ is a singular multivariate normal density, and for all l such that $\omega_l = 0$, $y_l^*|\mathbf{z}$ is a degenerate distribution. In this case, the IRT formulation of the UT model is given by (16) with (17) for all l such that $\omega_l \neq 0$, and by (16) with

$$\int_{R_l} \phi\left(y_l^* | \mathbf{z} : \mathbf{k}_l'\mathbf{z} + \mu_{y_l^*}, \omega_l\right) dy_l^* = (I_z)^{y_l}(1 - I_z)^{1-y_l} \quad \text{where} \quad I_z = \begin{cases} 1 & \text{if} \quad \mathbf{k}_l'\mathbf{z} + \mu_{y_l^*} > 0 \\ 0 & \text{if} \quad \mathbf{k}_l'\mathbf{z} + \mu_{y_l^*} < 0 \end{cases} \tag{19}$$

for all l such that $\omega_l = 0$.

Whenever $\omega_l = 0$ one of the two stimuli in pair l is consistently chosen for fixed $\mathbf{z}$. To see why in this case intransitivities may still arise, consider three objects $\{o_1, o_2, o_3\}$. Assume $\omega_{(1,2)} > 0, \omega_{(1,3)} > 0, \omega_{(2,3)} = 0$. This implies that subjects consistently choose between objects 2 and 3, but intransitivities may arise as subjects need not consistently choose between objects 1 and 2 or between objects 1 and 3.

In some applications researchers may have some prior substantive knowledge to believe that intransitivities should occur only for certain combinations of objects, thus leading to models with more than $n!$ but less than $2^{\tilde{n}}$ expected patterns. We shall refer to these models as *models with localized intransitivities*. Within a Thurstonian framework, such models with localized intransitivities arise when (a) some but not all elements of $\boldsymbol{\omega}$ are zero and (b) $\mathbf{P}_{y^*}$ is not positive definite. It can be readily verified that when these two conditions are satisfied, intransitive paired comparisons patterns that except for the columns in which $\omega_l \neq 0$ equal a transitive pattern are assigned zero probability. Note that since when $n = 3$ $\mathbf{P}_{y^*}$ is positive definite whenever any element of $\boldsymbol{\omega}$ is nonzero, models with localized intransitivities only exists for $n \geq 4$.

7. Model Estimation and Testing

Using the Lemma in section 4, any of the methods proposed for the estimation of the normal ogive model can be directly applied to estimate Thurstonian models for paired comparisons data: (a) limited information methods (e.g., Christoffersson, 1975; Muthén, 1978, 1984, 1993; Küsters, 1987), (b) full information methods based on the EM algorithm (Bock & Aitkin, 1981), or (c) resampling methods (e.g., Albert, 1992; Hajivassiliou, 1993; Schilling, 1993). Furthermore, any available computer program for the estimation of the normal ogive model can be used to estimate these models provided that the Thurstonian modeling constraints can be imposed on the thresholds and tetrachoric correlations of the normal ogive model via (12).

As paired comparisons experiments are rather time consuming, only small samples are generally collected in applications of this methodology. As a result, contingency tables, which are of size $2^{\tilde{n}}$, will generally be very sparse. This may give an edge to limited information estimators over full information estimators, as the former only make use of the lower order marginals of the contingency table which may be well estimated even in sparse tables. Furthermore, limited information estimators are considerably faster than full information estimators.

In this paper we shall employ three-stage limited information estimators proposed by Muthén (1978, 1993; Muthén, du Toit & Spisic, in press). Let $\boldsymbol{\kappa} = (\boldsymbol{\tau}, \boldsymbol{\rho})'$ where $\boldsymbol{\tau} = -\boldsymbol{\mu}_{y^*}$ and now $\boldsymbol{\rho}$ is used to denote the lower diagonal elements of $\mathbf{P}_{y^*}$. In a first stage each element of $\boldsymbol{\tau}$ is estimated separately as

$$\hat{\tau}_l = -\Phi_1^{-1}(\dot{p}_l), \qquad l = 1, \ldots, \tilde{n} \tag{20}$$

where $\dot{p}_l$ is the sample counterpart of $\dot{\pi}_l = \Pr(y_l = 1)$.

In the second stage, each element of $\mathbf{P}_{y^*}$ is estimated separately given the first stage estimates as

$$\hat{\rho}_{ll'} = \Phi_2^{-1}\big(\dot{p}_{ll'} | -\hat{\tau}_l, -\hat{\tau}_{l'}\big), \qquad l = 2, \ldots, \tilde{n}; l' = 1, \ldots, l-1 \tag{21}$$

where $\dot{p}_{ll'}$ is the sample counterpart of $\dot{\pi}_{ll'} = \Pr(y_l = 1, y_{l'} = 1)$.

Finally, in a third stage the parameters of any Thurstonian model collected in a q-dimensional vector $\boldsymbol{\theta}$ are estimated by minimizing

$$F = (\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\boldsymbol{\theta}))'\hat{\mathbf{W}}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\boldsymbol{\theta})) \tag{22}$$

where $\hat{\mathbf{W}}$ is a matrix converging in probability to $\mathbf{W}$, a positive definite matrix.

Let $\tilde{\mathbf{p}}_2 = (\dot{p}_1, \cdots, \dot{p}_{\tilde{n}}, \dot{p}_{2,1}, \cdots, \dot{p}_{\tilde{n},\tilde{n}-1})'$ and $\tilde{\boldsymbol{\pi}}_2$ be the vector of its corresponding probabilities, where

$$\sqrt{N}(\tilde{\mathbf{p}}_2 - \tilde{\boldsymbol{\pi}}_2) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}) \tag{23}$$

and $\xrightarrow{d}$ denotes convergence in distribution. Muthén (1978) showed that

$$\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}) \stackrel{a}{=} \tilde{\boldsymbol{\Delta}}^{-1}\sqrt{N}(\tilde{\mathbf{p}}_2 - \tilde{\boldsymbol{\pi}}_2) \tag{24}$$

where $\tilde{\boldsymbol{\Delta}} = \frac{\partial \tilde{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\kappa}'}$, and $\stackrel{a}{=}$ denotes asymptotic equality. Thus,

$$\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Xi} = \tilde{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\tilde{\boldsymbol{\Delta}}^{-1'}). \tag{25}$$

Therefore, obvious choices of $\hat{\mathbf{W}}$ in (22) are $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$ (Muthén, 1978), $\hat{\mathbf{W}} = \text{diag}(\hat{\boldsymbol{\Xi}})^{-1}$ (Muthén, du Toit & Spisic, in press), and $\hat{\mathbf{W}} = \mathbf{I}$ (Muthén, 1993). Using standard results for weighted least squares estimators (e.g., Browne, 1984; Satorra, 1989; Satorra & Bentler, 1994)

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \stackrel{a}{=} \mathbf{H}\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}) \qquad \mathbf{H} = (\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W} \tag{26}$$

$$N\hat{F} \xrightarrow{d} \textstyle\sum_{i=1}^{r} \alpha_i \chi_i^2 \tag{27}$$

where $\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}'}$, and $r = \frac{\tilde{n}(\tilde{n}+1)}{2} - q$ is the number of degrees of freedom of the model. In (27), the χ_i^2's are independent chi-square variables with one degree of freedom and the α_i's are the r nonnull eigenvalues of

$$\mathbf{M} = \mathbf{W}(\mathbf{I} - \boldsymbol{\Delta}\mathbf{H})\boldsymbol{\Xi}. \tag{28}$$

It then follows (Muthén, 1993) that

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{H}\boldsymbol{\Xi}\mathbf{H}'). \tag{29}$$

Now, when $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$, (27) and (29) simplify to $N\hat{F} \xrightarrow{d} \chi^2_r$, and $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, (\boldsymbol{\Delta}'\boldsymbol{\Xi}^{-1}\boldsymbol{\Delta})^{-1})$, respectively, and we obtain an estimator that asymptotically has minimum variance among the class of estimators based on the first and second order marginals of the contingency table. On the other hand, when $\hat{\mathbf{W}} = \text{diag}(\hat{\boldsymbol{\Xi}})^{-1}$ or $\hat{\mathbf{W}} = \mathbf{I}$, a goodness of fit of the model can be obtained (Muthén, 1993; Muthén et al., in press) by scaling $T := N\hat{F}$ by its mean or adjusting it by its mean and variance so that it approximates a chi-square distribution as follows

$$T_s = \frac{r}{\text{Tr}[\mathbf{M}]}T \qquad T_a = \frac{\text{Tr}[\mathbf{M}]}{\text{Tr}[\mathbf{M}^2]}T \tag{30}$$

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 $d = \frac{(\text{Tr}[\mathbf{M}])^2}{\text{Tr}[\mathbf{M}^2]}$ degrees of freedom.

These statistics can be used to test $H_0 : \boldsymbol{\kappa} = \boldsymbol{\kappa}(\boldsymbol{\theta})$. However, as discussed by Muthén (1993), verifying this hypothesis is only meaningful if the distributional hypothesis of dichotomized multivariate normality is correct, but a test of this hypothesis is only currently available for triplets of binary variables (Muthén & Hofacker, 1988).

Rather than employing this approach, we propose testing directly $H_0 : \tilde{\boldsymbol{\pi}}_2 = \tilde{\boldsymbol{\pi}}_2(\boldsymbol{\theta})$. By a Taylor series expansion

$$\tilde{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}}) \overset{a}{=} \tilde{\boldsymbol{\pi}}_2(\boldsymbol{\theta}) + \frac{\partial \tilde{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \tilde{\boldsymbol{\pi}}_2 + \tilde{\boldsymbol{\Delta}}\boldsymbol{\Delta}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}).$$

Hence, by (24) and (26)

$$\sqrt{N}(\tilde{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}}) - \tilde{\boldsymbol{\pi}}_2) \overset{a}{=} \tilde{\boldsymbol{\Delta}}\boldsymbol{\Delta}\mathbf{H}\tilde{\boldsymbol{\Delta}}^{-1}\sqrt{N}(\tilde{\mathbf{p}}_2 - \tilde{\boldsymbol{\pi}}_2). \tag{31}$$

Now, since

$$\sqrt{N}(\tilde{\mathbf{p}}_2 - \tilde{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}})) = \sqrt{N}(\tilde{\mathbf{p}}_2 - \tilde{\boldsymbol{\pi}}_2) - \sqrt{N}(\tilde{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}}) - \tilde{\boldsymbol{\pi}}_2) \overset{a}{=} \sqrt{N}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\boldsymbol{\Delta}\mathbf{H}\tilde{\boldsymbol{\Delta}}^{-1})(\tilde{\mathbf{p}}_2 - \tilde{\boldsymbol{\pi}}_2),$$

we obtain from (23)

$$\sqrt{N}\hat{\tilde{\mathbf{e}}} := \sqrt{N}(\tilde{\mathbf{p}}_2 - \tilde{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}})) \xrightarrow{d} N(\mathbf{0}, \tilde{\mathbf{M}}) \tag{32}$$

$$\tilde{\mathbf{M}} = (\mathbf{I} - \tilde{\boldsymbol{\Delta}}\boldsymbol{\Delta}\mathbf{H}\tilde{\boldsymbol{\Delta}}^{-1})\boldsymbol{\Gamma}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\boldsymbol{\Delta}\mathbf{H}\tilde{\boldsymbol{\Delta}}^{-1})'. \tag{33}$$

Consider now the test statistic $\tilde{T} = N\hat{\tilde{\mathbf{e}}}'\hat{\tilde{\mathbf{e}}}$. Since $\tilde{T} \xrightarrow{d} \sum_{i=1}^{r} \tilde{\alpha}_i \chi_i^2$, where the $\tilde{\alpha}_i$'s are the r nonnull eigenvalues of $\tilde{\mathbf{M}}$ (Box, 1954, Theorem 2.1). To test $H_0 : \tilde{\boldsymbol{\pi}}_2 = \tilde{\boldsymbol{\pi}}_2(\boldsymbol{\theta})$, following Satorra and Bentler (1994), we propose scaling $\tilde{T}$ by its mean or adjusting it by its mean and variance so that it approximates a chi-square distribution using

$$\tilde{T}_s = \frac{r}{\text{Tr}[\tilde{\mathbf{M}}]}\tilde{T} \qquad \tilde{T}_a = \frac{\text{Tr}[\tilde{\mathbf{M}}]}{\text{Tr}[\tilde{\mathbf{M}}^2]}\tilde{T}. \tag{34}$$

$\tilde{T}_s$ is to be referred to a chi-square distribution with r degrees of freedom, whereas $\tilde{T}_a$ is to be referred to a chi-square distribution with $\tilde{d} = \frac{(\text{Tr}[\tilde{\mathbf{M}}])^2}{\text{Tr}[\tilde{\mathbf{M}}^2]}$ degrees of freedom.

It is interesting to point out that in the case of the Case III, Case V and UT models, the reduced form parameters of the model $\boldsymbol{\kappa}$ can be expressed as a linear function of the model parameters, say $\boldsymbol{\kappa} = \boldsymbol{\Delta}\boldsymbol{\theta} + \mathbf{c}$. As a result, for these models (22) has a closed form solution

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{\Delta}'\hat{\mathbf{W}}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\hat{\mathbf{W}}(\hat{\boldsymbol{\kappa}} - \mathbf{c}) \tag{35}$$

Should an improper solution arise in these models with closed form solutions since zero elements in $\boldsymbol{\omega}$ are admissible, a boundary solution can be obtained by sequentially employing (35), setting the smallest improper $\rho_{ii'}$ equal to its boundary value until a proper solution is obtained. This is equivalent to the method of successive averaging (e.g., Hubert & Arabie, 1995) and it is more efficient computationally than employing reparameterization techniques (e.g., McDonald, 1980). Note, however, that standard errors and goodness of fit tests obtained in boundary solutions would be incorrect (Shapiro, 1985, 1988). Perhaps these could be obtained along the lines of Shapiro (1985, 1988) or employing resampling methods as suggested by Dijkstra (1992).

In closing this section, it should be pointed out that when a model assigns zero probabilities to some binary patterns (as in the case of models with localized intransitivities, or in the case of ranking data), the degrees of freedom available for testing will be reduced. Let $\mathbf{Y}^{\bullet}$ be a $t \times \tilde{n}$ matrix ($t \leq 2^{\tilde{n}}$) containing all binary patterns with nonzero model probabilities. Then, following Maydeu-Olivares (1999, Appendix 2) the number of mathematically independent elements in $\tilde{\boldsymbol{\pi}}_2$ (and hence in $\boldsymbol{\kappa}$) equals the rank of $\mathbf{T} = \begin{pmatrix}\mathbf{T}_1\\ \mathbf{T}_2\end{pmatrix}$, where $\mathbf{T}_1 = \mathbf{Y}^{\bullet\prime}$, and the kth row of $\mathbf{T}_2$ is given by

$$\mathbf{t}'_k = \mathbf{y}_l^{\bullet\prime} \odot \mathbf{y}_{l'}^{\bullet\prime}, \qquad l = 2, \ldots, \tilde{n}; l' = 1, \ldots, l-1 \tag{36}$$

where $\mathbf{y}_l^{\bullet}$ denotes the lth column of $\mathbf{Y}^{\bullet}$, and $\odot$ denotes an elementwise (Hadamard) product. Hence the number of degrees of freedom available for testing models with localized intransitivities equals $\text{rank}(\mathbf{T}) - q$.

8. Simulation Study

The asymptotically optimal WLS estimator has been repeatedly shown to behave poorly in small samples, particularly in large models (e.g., Muthén, 1993, Reboussin & Liang, 1998). Thus, recently, interest has turned to estimators such as DWLS with $\hat{\mathbf{W}} = \text{diag}(\hat{\boldsymbol{\Xi}})^{-1}$ (Muthén, du Toit & Spisic, in press), and ULS with $\hat{\mathbf{W}} = \mathbf{I}$ (Muthén, 1993), which do not require the inversion of the large symmetric matrix $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$.

No published study has compared the actual performance of the DWLS and ULS approaches. We shall therefore perform here a small simulation study to compare these approaches in estimating an unrestricted Thurstonian model for paired comparisons data. The simulation study will also allow us to investigate the small sample behavior of the overall goodness of fit tests just introduced.

Two models will be considered, a small and a medium size model with $n = 4$ and $n = 7$ objects respectively. The parameters used to generate the data were

$$n = 4 \text{ objects}$$

$$\mu_t = \begin{pmatrix} 0.5\\ 0\\ -0.5\\ 0\end{pmatrix} \quad \mathbf{P}_t = \begin{pmatrix} 1 & & & \\ 0.8 & 1 & & \\ 0.7 & 0.6 & 1 & \\ 0.8 & 0.7 & 0.6 & 1\end{pmatrix}$$

$$n = 7 \text{ objects}$$

$$\mu_t = \begin{pmatrix} 0.5 \\ 0 \\ -0.5 \\ 0 \\ 0.5 \\ -0.5 \\ 0 \end{pmatrix} \quad \mathbf{P}_t = \begin{pmatrix} 1 & & & & & & \\ 0.8 & 1 & & & & & \\ 0.7 & 0.6 & 1 & & & & \\ 0.8 & 0.7 & 0.6 & 1 & & & \\ 0.8 & 0.7 & 0.6 & 0.8 & 1 & & \\ 0.7 & 0.6 & 0.8 & 0.7 & 0.6 & 1 & \\ 0.8 & 0.7 & 0.6 & 0.8 & 0.7 & 0.6 & 1 \end{pmatrix}$$

Two sample sizes will be considered, $N = 300$ and $N = 100$. 1000 replicates were used in each experimental cell. To increase the comparability of the results the same starting seed was used within each cell. The parameter estimates and estimated standard errors for the small model are shown in Table 1, and the goodness of fit statistics are shown in Tables 2 and 3.

For this model, we also show the results obtained using the full WLS estimator. In this case, we used a generalized inverse of the weight matrix whenever this was found to be singular.

We shall define estimation bias as $\frac{\overline{x}_{\hat{\theta}} - \theta_0}{\theta_0}$, standard error bias as $\frac{\overline{x}_{SE(\hat{\theta})} - sd_{\hat{\theta}}}{sd_{\hat{\theta}}}$, and coverage rate as the percentage of estimated 95% confidence intervals ($\hat{\theta} \pm 1.96\ SE(\hat{\theta})$) that included the true parameter. We note in Table 1 that both the DWLS and WLS correlation parameter estimates show a downward bias. This increases in the smaller sample size and it is larger for WLS than for DWLS. We also note that the estimated standard errors are largest for ULS, followed by DWLS, and then by WLS. The estimator that shows the smallest standard error bias is ULS. However, the estimation and standard error biases for DWLS are small. Turning to the results for the tests of the structural restrictions shown in Table 2, we see that both the mean (T_s) and mean and variance (T_a) corrected statistics perform reasonably at $N = 300$, but when $N = 100$, T_a clearly outperforms T_s, the latter being too optimistic. This is true for both ULS and DWLS estimation. The test statistic associated with WLS, $N\hat{F}$, is too optimistic at $N = 300$ and clearly inadequate at $N = 100$. Finally, we see in Table 3 that the overall $\tilde{T}_a$ statistic works well for all three estimators even at $N = 100$, whereas the $\tilde{T}_s$ is too liberal. In sum, we conclude that even with the small model considered here, WLS is clearly inadequate at the small sample sizes considered: Its parameter estimates and standard errors for the correlation structure show a large downward bias. The differences obtained when using DWLS and ULS are small. ULS parameter estimates, however, do not show the consistent downward bias of DWLS parameter estimates, and its standard errors show a smaller bias.

Tables 4 and 5 summarize the results corresponding to the larger model (7 objects). Here the number of parameters is rather large and the results shown in Table 4 are pooled across parameter estimates having the same true value. Again, we observe that both the DWLS and ULS estimators work well even when $N = 100$, and that the ULS estimator shows a better performance than the DWLS estimator, as the DWLS correlation parameter estimates show a downward bias, whereas the ULS correlation estimates show almost no bias. Furthermore, the standard error bias for the correlation parameters is smaller for ULS than for DWLS. We also observe in Table 5 than it is inappropriate to use the T_s and $\tilde{T}_s$ statistics for this medium size model. On the other hand, the T_a and $\tilde{T}_a$ statistics work well in all cases.

9. Application

We investigated compact car preferences among college students. Using a multiple judgment paired comparisons design, 289 subjects were asked which car they would buy if they could afford one. In this example we shall analyze their responses to these four cars {Opel Corsa, Renault Clio, Seat Ibiza, Volkswagen Polo}. With four objects, there are $2^6 = 64$ possible paired comparisons patterns, of which $4! = 24$ are transitive patterns. However, only 41 distinct patterns

TABLE 1.
Parameter estimates and estimated standard errors for an unrestricted model for 4 objects

		$N = 300$									$N = 100$								
		ULS			DWLS			WLS			ULS			DWLS			WLS		
par	true	$\overline{x}$ est.	$\overline{x}$ SEs	sd est	$\overline{x}$ est.	$\overline{x}$ SEs	sd est	$\overline{x}$ est.	$\overline{x}$ SEs	sd est	$\overline{x}$ est.	$\overline{x}$ SEs	sd est	$\overline{x}$ est.	$\overline{x}$ SEs	sd est	$\overline{x}$ est.	$\overline{x}$ SEs	sd est
μ_1	0.50	0.502	0.059	0.058	0.501	0.059	0.058	0.500	0.058	0.058	0.504	0.102	0.101	0.501	0.102	0.100	0.499	0.097	0.109
μ_2	0	0.000	0.061	0.059	−0.000	0.061	0.059	−0.001	0.060	0.060	−0.003	0.106	0.105	−0.003	0.106	0.105	−0.005	0.100	0.110
μ_3	−0.50	−0.500	0.068	0.065	−0.499	0.067	0.065	−0.499	0.066	0.065	−0.504	0.118	0.115	−0.502	0.117	0.113	−0.508	0.111	0.121
ρ_{21}	0.80	0.802	0.043	0.042	0.799	0.043	0.043	0.792	0.041	0.044	0.802	0.073	0.075	0.794	0.072	0.077	0.771	0.064	0.088
ρ_{31}	0.70	0.702	0.046	0.045	0.698	0.046	0.045	0.686	0.043	0.048	0.708	0.080	0.083	0.694	0.079	0.084	0.661	0.066	0.101
ρ_{32}	0.60	0.599	0.046	0.048	0.596	0.045	0.047	0.587	0.040	0.044	0.597	0.079	0.080	0.585	0.076	0.080	0.565	0.062	0.082
ρ_{41}	0.80	0.802	0.042	0.043	0.799	0.042	0.043	0.792	0.041	0.045	0.803	0.073	0.075	0.795	0.072	0.077	0.771	0.064	0.087
ρ_{42}	0.70	0.702	0.042	0.044	0.699	0.042	0.043	0.694	0.039	0.043	0.701	0.073	0.075	0.692	0.071	0.076	0.676	0.062	0.083
ρ_{43}	0.60	0.601	0.046	0.046	0.597	0.045	0.045	0.589	0.040	0.043	0.601	0.079	0.082	0.589	0.076	0.081	0.569	0.062	0.084

Notes: 1000 replications; 6 binary variables are modeled; $\mu_4 = 0$ for identification purposes; 12 *df*.

TABLE 2.
Tests of the structural restrictions for an unrestricted model for 4 objects

		$N = 300$					$N = 100$				
$H_0 : \boldsymbol{\kappa} = \boldsymbol{\kappa}(\boldsymbol{\theta})$		ULS		DWLS		WLS	ULS		DWLS		WLS
	stat	T_s	T_a	T_s	T_a	NF	T_s	T_a	T_s	T_a	NF
	Mean	12.2	9.58	12.2	10.3	13.2	12.9	9.1	12.9	9.7	16.0
	Var.	29.5	18.03	27.9	19.5	32.5	41.9	20.1	33.9	18.3	60.7
	1%	2.5	1.6	2.0	1.4	2.6	3.5	1.4	2.7	1.4	10.3
Nominal	5%	5.7	4.0	5.2	4.2	8.6	9.4	5.8	8.8	6.0	21.3
Rates	10%	11.1	8.7	11.0	8.9	16.1	14.2	11.6	15.6	12.9	30.3
	20%	21.8	20.1	22.2	21.1	27.1	24.7	22.9	25.3	23.8	43.5

Notes: 1000 replications; 6 binary variables are modeled; 12 *df*.

TABLE 3.
Overall limited information tests for an unrestricted model for 4 objects

		$N = 300$						$N = 100$					
$H_0 : \tilde{\boldsymbol{\pi}}_2 = \tilde{\boldsymbol{\pi}}_2(\boldsymbol{\theta})$		ULS		DWLS		WLS		ULS		DWLS		WLS	
	stat	$\tilde{T}_s$	$\tilde{T}_a$	$\tilde{T}_s$	$\tilde{T}_a$	$\tilde{T}_s$	$\tilde{T}_a$	$\tilde{T}_s$	$\tilde{T}_a$	$\tilde{T}_s$	$\tilde{T}_a$	$\tilde{T}_s$	$\tilde{T}_a$
	Mean	12.3	4.0	11.9	3.9	12.1	4.3	12.5	4.0	12.1	3.8	13.5	5.2
	Var.	74.1	8.0	70.0	7.6	65.9	8.4	79.8	8.0	74.6	7.6	80.5	12.0
	1%	7.5	1.2	7.2	1.0	6.5	1.2	8.7	1.3	7.3	1.0	9.5	1.9
Nominal	5%	13.8	5.7	12.8	5.2	13.2	5.2	14.2	6.1	13.0	5.3	16.0	8.4
Rates	10%	18.7	10.6	17.7	9.6	17.2	10.0	19.2	11.2	17.9	10.4	21.2	13.9
	20%	26.5	20.0	24.4	18.9	25.9	18.6	26.1	19.8	24.2	19.3	30.1	23.2

Notes: 1000 replications; 6 binary variables are modeled; 12 *df*.

were observed in this sample, as 267 subjects (92% of the sample) yielded transitive patterns (all transitive patterns were observed). The paired comparisons patterns and their observed frequencies in this sample are given in Table 6.

It was shown in the previous section that the results obtained using the ULS or DWLS were very similar, whereas it is inappropriate to use full WLS. For all the models estimated here, the parameter estimates, estimated standard errors and p-values associated with the goodness of fit tests obtained using ULS and DWLS agree at least to two significant digits. Hence, since we have seen in the previous section that the ULS estimator has a slight better performance than the DWLS estimator, only the ULS results will be reported here.

The ULS parameter estimates and estimated asymptotic standard errors for the unrestricted model are shown in Table 7. As can be seen in this table, all elements of $\boldsymbol{\Omega}$ except for ω_1 are significantly equal to zero (or equivalently all elements of $\mathbf{P}_t$ but ρ_{21} are significantly equal to 0.5). This was expected since there are so few intransitive observations in this sample.

The goodness of fit statistics for this model are shown in Table 8. In this table, the usefulness of the proposed test of the model to the first and second order marginals of the contingency table is readily apparent. The statistics for testing the restrictions introduced by the UT model on the reduced form parameters (thresholds and tetrachoric correlations) suggests that the UT model fits very well these data. Yet, these statistics are misleading. The $\tilde{T}_s$ and $\tilde{T}_a$ statistics reveal that the UT model does not fit well the first and second order marginals of the contingency table.

TABLE 4.
Parameter estimates and estimated standard errors for an unrestricted model for 7 objects

	ULS						DWLS					
	$N = 300$											
par	$\bar{x}$ est.	$\bar{x}$ SEs	sd est.	bias est.	bias SEs	cover-age	$\bar{x}$ est.	$\bar{x}$ SEs	sd est.	bias est.	bias SEs	cover-age
$\mu = 0.5$	0.500	0.053	0.053	0.001	0.013	0.959	0.499	0.053	0.052	−0.002	0.017	0.959
$\mu = 0$	−0.001	0.052	0.051	—	0.007	0.951	−0.001	0.052	0.051	—	0.009	0.954
$\mu = -0.5$	−0.505	0.062	0.062	0.009	0.012	0.954	−0.504	0.062	0.061	0.007	0.013	0.952
$\rho = 0.8$	0.800	0.029	0.029	−0.001	−0.021	0.941	0.796	0.029	0.030	−0.005	−0.043	0.939
$\rho = 0.7$	0.702	0.033	0.033	0.003	−0.006	0.944	0.696	0.033	0.033	−0.006	−0.024	0.940
$\rho = 0.6$	0.601	0.037	0.037	0.002	0.005	0.948	0.594	0.035	0.036	−0.010	−0.017	0.945
	$N = 100$											
$\mu = 0.5$	0.504	0.093	0.094	0.008	−0.015	0.950	0.500	0.092	0.093	−0.001	−0.008	0.950
$\mu = 0$	−0.001	0.090	0.090	—	−0.006	0.949	−0.001	0.090	0.090	—	0.002	0.950
$\mu = -0.5$	−0.510	0.108	0.111	0.021	−0.030	0.943	−0.507	0.108	0.110	0.014	−0.022	0.941
$\rho = 0.8$	0.801	0.049	0.052	0.001	−0.055	0.922	0.789	0.048	0.055	−0.014	−0.117	0.914
$\rho = 0.7$	0.704	0.057	0.059	0.006	−0.043	0.925	0.686	0.055	0.061	−0.021	−0.099	0.915
$\rho = 0.6$	0.604	0.063	0.065	0.006	−0.025	0.928	0.581	0.059	0.064	−0.031	−0.084	0.913

Notes: 1000 replications; 21 binary variables are modeled; 204 *df*.

TABLE 5.
Goodness of fit tests for an unrestricted model for 7 objects

		$N = 300$				$N = 100$			
		ULS		DWLS		ULS		DWLS	
	stat	T_s	T_a	T_s	T_a	T_s	T_a	T_s	T_a
$H_0 : \boldsymbol{\kappa} = \boldsymbol{\kappa}(\boldsymbol{\theta})$	Mean	208.3	84.0	211.9	90.3	216.5	46.9	223.9	52.2
	Var.	658.2	104.1	670.9	116.4	1019.8	49.6	1016.3	56.8
	1%	12.1	0.7	16.1	0.9	22.3	0.7	30.1	0.5
Nominal	5%	55.8	2.9	62.3	4.1	63.1	3.8	73.5	3.9
Rates	10%	55.8	7.2	62.3	10.3	63.1	7.3	73.5	11.7
	20%	55.8	18.6	62.3	24.0	63.1	20.5	73.5	28.8
$H_0 : \tilde{\boldsymbol{\pi}}_2 = \tilde{\boldsymbol{\pi}}_2(\boldsymbol{\theta})$	Mean	208.5	23.9	205.7	23.7	209.2	20.3	206.5	20.2
	Var.	3347.3	45.11	3289.8	44.2	3406.0	32.2	3325.6	31.9
	1%	27.8	1.0	26.3	0.7	28.0	1.1	26.2	1.2
Nominal	5%	48.1	6.4	45.9	6.2	49.6	4.2	47.5	3.6
Rates	10%	48.1	11.0	45.9	9.9	49.6	9.3	47.5	8.5
	20%	48.1	22.3	45.9	20.6	49.6	19.3	47.5	18.0

Notes: 1000 replications; 21 binary variables are modeled; 204 *df*.

For this small model, we also calculated a full information X^2 statistic where the expected probabilities for each paired comparison pattern were computed using the IRT formulation of the model. It remains to be shown that a X^2 statistic has an asymptotic chi-square distribution with $2^{\tilde{n}} - q - 1$ degrees of freedom when model parameters are estimated using the limited information procedures employed here. Therefore, the degrees of freedom and associated p-value for the X^2 statistic reported in Table 8 are provided for illustrative purposes only. Under the assumption that they are asymptotically correct, they would suggest that the UT model does not fit the paired comparisons patterns adequately. An examination of the expected frequencies shown in Table 6 reveals that the model reproduces adequately all the observed patterns except for the intransitive patterns 30 and 54. According to the model, these patterns are very unlikely to be observed, but had a single occurrence in this sample. The standardized residuals for these patterns are 4.306 and 5.560, respectively, and thus these two observations account for over 60% of the value of the X^2 statistic.

It may be that the responses of these two subjects were due to carelessness during the experimental procedure. For illustrative purposes we shall remove these two observations from the data and re-estimate the model as when they are removed an improper solution is obtained. We obtained a boundary solution by applying (35) sequentially, setting the smallest improper $\rho_{ii'}$ equal to $\frac{1}{2}$ until a proper solution was obtained. The parameter estimates and standard errors for this boundary solution are also shown in Tables 7 and 8. The standard errors and goodness of fit tests are incorrect and are only shown for illustrative purposes. Not surprisingly, the parameter estimates for the boundary solution are almost identical to those obtained from the complete data, but now the overall limited information statistic and the full information statistic suggest that the model fits well these data after the two outlier observations are removed.

Alternatively, given the parameter estimates and standard errors for the unrestricted model, one may consider fitting a model with these restrictions $\{\mu_4 = 0; \mu_3 = \mu_2; \rho_{31} = \rho_{32} = \rho_{41} = \rho_{42} = \rho_{43} = \frac{1}{2}\}$. It turns out that this is a model with localized intransitivities as $\omega_1 > 0$ and $\mathbf{P}_{y^*}$ is of deficient rank (its rank equals 4). According to this model, 28 of the intransitive patterns have zero probability. However, seven of these patterns {6, 9, 24, 30, 41, 50, 54} were observed in the complete data, each having a single occurrence. In other words according to the model these

TABLE 6.
Observed and expected frequencies of paired comparisons patterns for the compact cars data

no.	pattern	obs.	unrestrict.	restrict.	no.	pattern	obs.	unrestrict.	restrict.
∗ 1	111111	15	15.795	16.676	∗33	011111	10	11.331	11.567
∗ 2	111110	24	17.520	19.527	∗34	011110	12	11.793	13.606
3	111101	0	0.657	0	35	011101	0	0.158	0
∗ 4	111100	13	19.438	17.920	36	011100	2	2.401	1.597
∗ 5	111011	12	12.657	13.541	37	011011	1	2.030	1.265
6	111010	1	0.777	0	38	011010	0	0.221	0
∗ 7	111001	10	11.729	13.684	39	011001	0	0.434	0.394
∗ 8	111000	20	16.884	15.424	40	011000	1	0.823	0.415
9	110111	1	0.209	0	41	010111	1	0.440	0
10	110110	1	1.737	1.058	∗42	010110	6	6.386	7.546
11	110101	0	0.119	0	43	010101	0	0.047	0
∗12	110100	18	17.036	16.164	∗44	010100	8	10.051	12.333
13	110011	0	0.124	0	45	010011	0	0.062	0
14	110010	0	0.084	0	46	010010	0	0.123	0
15	110001	0	0.708	0	47	010001	0	0.134	0
∗16	110000	19	16.681	14.814	48	010000	2	2.184	1.382
17	101111	1	1.326	0.858	∗49	001111	6	6.220	5.840
18	101110	0	0.050	0	50	001110	1	0.126	0
19	101101	0	0.109	0	51	001101	0	0.122	0
20	101100	0	0.079	0	52	001100	0	0.319	0
∗21	101011	15	10.702	10.546	∗53	001011	10	8.817	8.186
22	101010	0	0.013	0	•54	001010	1	0.030	0
∗23	101001	8	10.544	11.183	55	001001	0	1.791	1.087
24	101000	1	0.457	0	56	001000	1	0.191	0
25	100111	0	0.442	0.197	∗57	000111	4	4.421	3.948
26	100110	0	0.208	0.232	∗58	000110	3	3.269	4.491
27	100101	0	0.178	0	59	000101	0	0.210	0
28	100100	1	1.066	0.931	∗60	000100	8	5.199	6.358
29	100011	4	1.539	0.733	∗61	000011	6	6.010	4.794
•30	100010	1	0.049	0	62	000010	1	0.289	0
∗31	100001	12	12.068	11.776	∗63	000001	8	8.389	8.347
∗32	100000	13	11.368	13.647	∗64	000000	7	8.880	9.691

Notes: $N = 289$; ∗ transitive patterns, • outlier observations deleted when fitting the boundary model

seven observations should not have occurred. Note that the two observations that we removed in our previous analysis because of their large standardized residuals belong to this set. We fitted again this model with localized intransitivities to our data without these seven observations. The resulting parameter estimates and standard errors are also shown in Tables 7 and 8. The standard errors and goodness of fit indexes for this model are asymptotically correct (Dijkstra, 1992). In Table 6 we show the expected frequencies for each pattern. Both the limited and full information test suggest that this model matches very well the data (after removing seven observations). The rank of $\mathbf{T}$ for this model is 19 and therefore the number of degrees of freedom reported for this model in Table 8 are $19 - 3 = 16$.

10. Discussion and Conclusions

The method of paired comparisons is most often employed using a multiple judgment framework. Following Takane's (1987) seminal ideas, we have discussed how to estimate Thurstonian models for these data as a mean and correlation structure model with dichotomous indicators. Existing software can be used to estimate some of these models, such as

TABLE 7.
Estimated parameters and asymptotic standard errors

	unrestricted model		boundary model		restricted model	
par.	value	SE	value	SE	value	SE
μ_1	0.201	(0.066)	0.205	(0.066)	0.200	(0.067)
μ_2	−0.155	(0.068)	−0.156	(0.069)	−0.144	(0.061)
μ_3	−0.112	(0.069)	−0.115	(0.068)	−0.144	(0.061)
μ_4	0	*fixed*	0	*fixed*	0	*fixed*
ρ_{21}	0.658	(0.037)	0.655	(0.040)	0.633	(0.040)
ρ_{31}	0.502	(0.040)	0.5	*constrained*	0.5	*fixed*
ρ_{32}	0.556	(0.038)	0.552	(0.041)	0.5	*fixed*
ρ_{41}	0.561	(0.039)	0.556	(0.044)	0.5	*fixed*
ρ_{42}	0.503	(0.040)	0.5	*constrained*	0.5	*fixed*
ρ_{43}	0.504	(0.040)	0.5	*constrained*	0.5	*fixed*
ω_1	0.315	(0.073)	0.309	(0.079)	0.267	(0.080)
ω_2	0.004	(0.079)	0	*constrained*	0	*fixed*
ω_3	0.121	(0.078)	0.112	(0.088)	0	*fixed*
ω_4	0.113	(0.076)	0.103	(0.083)	0	*fixed*
ω_5	0.006	(0.080)	0	*constrained*	0	*fixed*
ω_6	0.009	(0.079)	0	*constrained*	0	*fixed*

Notes: $N = 289$ (unrestricted model), $N = 287$ (boundary model), $N = 282$ (restricted model); 1 = Opel Corsa, 2 = Renault Clio, 3 = Seat Ibiza, 4 = Volkswagen Polo; $\mu_4 = 0$ for identification purposes; μ_2 and μ_3 are constrained to be equal in the restricted model

TABLE 8.
Goodness of fit tests

	$H_0 : \boldsymbol{\kappa} = \boldsymbol{\kappa}(\boldsymbol{\theta})$			$H_0 : \tilde{\boldsymbol{\pi}}_2 = \tilde{\boldsymbol{\pi}}_2(\boldsymbol{\theta})$			$H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$		
Model	Statistic	*df*	*p*	statistic	*df*	*p*	statistic	*df*	*p*
Unrestricted	$T_s = 10.13$	12	0.61	$\tilde{T}_s = 9.00$	12	< 0.01	$X^2 = 81.65$	54	0.01
	$T_a = 7.82$	9.27	0.58	$\tilde{T}_a = 2.39$	4.40	0.01			
boundary	$T_s = 4.91$	12	0.96	$\tilde{T}_s = 4.51$	15	0.06	$X^2 = 38.02$	54	0.95
	$T_a = 2.92$	7.13	0.90	$\tilde{T}_a = 1.56$	5.94	0.12			
restricted	$T_s = 14.12$	16	0.59	$\tilde{T}_s = 19.70$	16	0.23	$X^2 = 24.12$	32	0.84
	$T_a = 8.08$	9.16	0.54	$\tilde{T}_a = 11.27$	4.28	0.27			

Notes: $N = 289$ (unrestricted model), $N = 287$ (boundary model), $N = 282$ (restricted model); $\boldsymbol{\pi}$ is used here to denote the paired comparison pattern probabilities (7); the *p*-values for the boundary model are incorrect and are shown only for illustrative purposes.

MPLUS (Muthén & Muthén, 1998) which implements the DWLS estimator. Note however, that in MPLUS the weight matrix of the reduced form parameters is computed as in Muthén (1984) instead of as in Muthén (1978) as we do here.

We have seen that when the sample thresholds and tetrachoric correlations can be expressed as a linear function of the Thurstonian model parameters both the ULS and DWLS estimators work well even for seven objects (21 dichotomous variables) and with sample sizes down to one hundred observations. Furthermore, we have proposed methods to assess the goodness of fit of the models to the first and second order marginals of the contingency table that overcome the single most serious problem faced when modeling these data, namely, how to assess the fit of the estimated models. The proposed tests work well under the conditions examined here. Further work is needed to investigate the behavior of the proposed estimators and tests in fitting

Thurstonian models that impose non-linear restrictions on the reduced form parameters (e.g., factor or ideal point models).

In closing, it should be noted that the ULS estimator employed here is closely related to the classical least squares procedures proposed for estimating Thurstonian models (see Arbuckle & Nugent, 1973; Bock & Jones, 1968, and references therein). However, these use first order information only (i.e., binary choices), which are assumed to be independent. Under multiple judgment sampling, this assumption is clearly violated. In contrast, the ULS estimator employed here uses first and second order information and takes into consideration the dependencies among the sample statistics to obtain asymptotically correct standard errors and goodness of fit tests.

References

Albert, J.H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, *17*, 251–269.

Arbuckle, L., & Nugent, J.H. (1973). A general procedure for parameter estimation for the law of comparative judgment. *British Journal of Mathematical and Statistical Psychology*, *26*, 240–260.

Bekker, P.A., Merckens, A., & Wansbeek, T.J. (1994). *Identification, equivalent models and computer algebra*. San Diego: Academic Press.

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

Bock, R.D., & Jones, L.V. (1968). *The measurement and prediction of judgment and choice*. San Francisco, CA: Holden-Day.

Böckenholt, U., & Dillon, W.R. (1997). Modeling within-subject dependencies in ordinal paired comparisons data. *Psychometrika*, *62*, 411–434.

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.

Brady, H.E. (1989). Factor and ideal point analysis for interpersonally incomparable data. *Psychometrika*, *54*, 181–202.

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

Carroll, J.D. (1980). Models and methods for multidimensional analysis of preference choice (or other dominance) data. In E.D. Lantermann & H. Feger (Eds.), *Similarity and choice*. Bern, Switzerland: Hans Huber.

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

De Soete, G., & Carroll, J.D. (1983). A maximum likelihood method for fitting the wandering vector model. *Psychometrika*, *48*, 553–566.

De Soete, G., Carroll, J.D., & DeSarbo, W.S. (1986). The wandering ideal point model: A probabilistic multidimensional unfolding model for paired comparison data. *Journal of Mathematical Psychology*, *30*, 28–41.

Dijkstra, T. K. (1992). On statistical inference with parameter estimates on the boundary of parameter space. *British Journal of Mathematical and Statistical Psychology*, *45*, 289–309.

Hajivassiliou, V.A. (1993). Simulation estimation methods for limited dependent variable models. In G.S. Maddala, C.R. Rao, & H.D. Vinod (Eds.), *Handbook of Statistics* (Vol. 11). New York, NY: Elsevier Science.

Heiser, W., & de Leeuw, J. (1981). Multidimensional mapping of preference data. *Mathématiques et Sciences Humaines*, *19*, 39–96.

Hubert, L., & Arabie, P. (1995). Iterative projection strategies for the least squares fitting of tree structures to proximity data. *British Journal of Mathematical and Statistical Psychology*, *48*, 281–317.

Küsters, U.L. (1987). *Hierarchische Mittelwert- und Kovarianztrukturmodelle mit nichtmetrischen endogenen Variablen* [Hierarchical mean and covariance structure models on nonmetric endogenous variables]. Heidelberg, Germany: Physica-Verlag.

Maydeu-Olivares, A. (1999). Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika*, *64*, 325–340.

McDonald, R.P. (1980). A simple comprehensive model for the analysis of covariance structures: Some remarks on applications. *British Journal of Mathematical and Statistical Psychology*, *33*, 161–183.

McDonald, R.P. (1985). *Factor analysis and related methods*. New York, NY: Lawrence Erlbaum.

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

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

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.

Muthén, L., & Muthén, B. (1998). *Mplus*. Los Angeles, CA: Muthén & Muthén.

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

Reboussin, B.A., & Liang, K.Y. (1998). An estimating equations approach for the LISCOMP model. *Psychometrika*, *63*, 165–182.

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 to developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Schilling, S. (1993). *Advances in full-information item factor analysis using the Gibbs sampler*. Unpublished doctoral dissertation, University of Chicago.

Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, *72*, 133–144.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, *56*, 49–62.

Takane, Y. (1980). Maximum likelihood estimation in the generalized case of Thurstone's model of comparative judgment. *Japanese Psychological Research*, *22*, 188–196.

Takane, Y. (1987). Analysis of covariance structures and probabilistic binary choice data. *Communication and Cognition*, *20*, 45–62.

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

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

Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review*, *79*, 281–299.

Manuscript received 21 SEP 1998
Final version received 6 OCT 1999