

ON THURSTONE'S MODEL FOR PAIRED COMPARISONS AND RANKING DATA

Albert Maydeu-Olivares

Dept. of Psychology. University of Barcelona. Paseo Valle de Hebrón, 171. 08035 Barcelona (Spain).

Summary. We investigate by means of a simulation study whether standard structural equation modeling techniques are suitable for estimating an unrestricted Thurstonian model to: (a) multiple judgment paired comparisons data, and (b) ranking data. We point out that Thurstone's (1927) original model is not a proper model for multiple judgment paired comparisons data as it assigns zero probability to all intransitive paired comparisons patterns. To fit multiple judgment paired comparisons data one must employ Takane's (1987) extension of Thurstone's original model, or alternatively, a Thurstonian correlation structure model, which we introduce. We found that for some models as few as 100 observations suffice to obtain accurate parameter estimates, standard errors and goodness of fit tests when 7 stimuli are modeled (21 binary variables). For other models, however, sample sizes of 1000 observations are needed. All in all, the procedure investigated appears to be an attractive choice to estimate these models.

Keywords: limited information, comparative data, random utility models.

1. Introduction

In 1927 Thurstone suggested a class of models for fitting paired comparison data that has been highly influential in the literature (see Bock & Jones, 1968). In 1931 Thurstone suggested his model could also be suitable for ranking data (after rankings are transformed into paired comparisons). Thurstone's model is simply a multivariate normal density with an structured mean vector and covariance matrix that has been dichotomized. Thus, it is somewhat natural to consider its estimation using existing procedures for structural equation modeling for dichotomous variables (Muthén, 1978, 1993).

In this paper we investigate how well an unrestricted Thurstonian model for paired comparisons and ranking data can be estimated using structural equation modeling procedures for dichotomous variables. To do so, we first introduce Thurstone's original model using matrix notation. Using this notation, it is easy to see that Thurstone's original model is a proper model for ranking data but it is not a proper model for multiple judgment paired comparisons data. Next, we

introduce two modifications of Thurstone's model that address this problem. We conclude our presentation with some simulation results to illustrate the performance of structural equation modeling procedures to estimate these models.

2. Thurstone's model

Thurstone (1927) introduced a model for paired comparisons characterized by three assumptions:

1. 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.
2. The stimulus whose value is larger at the moment of the comparison will be preferred by the subject.
3. These unobserved preferences are normally distributed in the population.

We assume each individual from a random sample responds to all possible $\tilde{n} = \frac{n(n-1)}{2}$ paired comparisons, where n denotes the number of stimuli being compared. This has been termed *multiple judgment sampling procedure* (Bock & Jones, 1968). Thus, we obtain a \tilde{n} -dimensional vector of binary variables \mathbf{y} from each respondent such that

$$y_{i,j} = \begin{cases} 1 & \text{if object } i \text{ is chosen} \\ 0 & \text{if object } j \text{ is chosen} \end{cases} \quad (1)$$

for all pairs $\{i, j\}$. Then, letting \mathbf{t} denote the n -dimensional vector of continuous preferences in Thurstone's model such that $\mathbf{t} \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$, we write

$$\mathbf{y}^* = \mathbf{A} \mathbf{t}, \quad (2)$$

where \mathbf{A} is a $\tilde{n} \times n$ design matrix where each column corresponds to one of the stimuli, and each row to one of the paired comparisons. For example, when $n = 4$, \mathbf{A} is

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}. \quad (3)$$

Finally,

$$y_{i,j} = \begin{cases} 1 & \text{if } y_{i,j}^* \geq 0 \\ 0 & \text{if } y_{i,j}^* < 0 \end{cases} \quad (4)$$

These equations describe in matrix notation Thurstone's (1927) model applied to multiple judgment sampling paired comparisons data. To analyze ranking data, Thurstone (1931) suggested transforming each ranking pattern into a vector of \tilde{n} binary variables using

$$y_{i,j} = \begin{cases} 1 & \text{if object } i \text{ is ranked above object } j \\ 0 & \text{if object } i \text{ is ranked below object } j \end{cases} \quad (5)$$

and applying his model for paired comparisons data to these data.

Under Thurstone's model the probability of any paired comparisons pattern is

$$\Pr\left(\bigcap_{i,j} y_{i,j}\right) = \int_{\mathbf{R}} \cdots \int_{\mathbf{R}} \phi_{\tilde{n}}(\mathbf{y}^* : \boldsymbol{\mu}_{y^*}, \boldsymbol{\Sigma}_{y^*}) d\mathbf{y}^* \quad (6)$$

$$\boldsymbol{\mu}_{y^*} = \mathbf{A}\boldsymbol{\mu}_t \quad \boldsymbol{\Sigma}_{y^*} = \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}' \quad (7)$$

where $\phi_{\tilde{n}}(\bullet)$ denotes a \tilde{n} -variate normal density function and \mathbf{R} is the multidimensional rectangular region formed by the product of intervals

$$R_{i,j} = \begin{cases} (0, \infty) & \text{if } y_{i,j} = 1 \\ (-\infty, 0) & \text{if } y_{i,j} = 0 \end{cases} \quad (8)$$

We can perform a change of variable of integration in (6) and standardize \mathbf{y}^* using

$$\mathbf{z}^* = \mathbf{D}(\mathbf{y}^* - \boldsymbol{\mu}_{y^*}) \quad \mathbf{D} = \left(\text{Diag}(\boldsymbol{\Sigma}_{y^*})\right)^{\frac{1}{2}} \quad (9)$$

As a result, $\boldsymbol{\mu}_{z^*} = \mathbf{0}$ and

$$\boldsymbol{\tau} = -\mathbf{D}\boldsymbol{\mu}_{y^*} = -\mathbf{D}\mathbf{A}\boldsymbol{\mu}_t \quad \mathbf{P}_{z^*} = \mathbf{D}\boldsymbol{\Sigma}_{y^*}\mathbf{D} = \mathbf{D}(\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}')\mathbf{D} \quad (10)$$

where $\boldsymbol{\tau}$ denotes a vector of thresholds $\tau_{i,j}$, and the off-diagonal elements of \mathbf{P}_{z^*} are tetrachoric correlations. As a result, (6), can be equivalently rewritten as

$$\Pr\left(\bigcap_{i,j} y_{i,j}\right) = \int_{\tilde{\mathbf{R}}} \cdots \int_{\tilde{\mathbf{R}}} \phi_{\tilde{n}}(\mathbf{z}^* : \mathbf{0}, \mathbf{P}_{z^*}) d\mathbf{z}^* \quad (11)$$

where $\tilde{\mathbf{R}}$ is now a multidimensional rectangular region formed by the product of intervals

$$\tilde{R}_{i,j} = \begin{cases} (\tau_{i,j}, \infty) & \text{if } y_{i,j} = 1 \\ (-\infty, \tau_{i,j}) & \text{if } y_{i,j} = 0 \end{cases}. \quad (12)$$

Equations (10), (11), and (12) define in fact a class of models as $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ can be restricted in various ways. Takane (1987) provides an excellent overview of restricted Thurstonian models. Here we shall concentrate on the *unrestricted* Thurstonian model, a model where only minimal identification restrictions are imposed on $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$.

A very interesting feature of Thurstone's original model is that since \mathbf{A} is of rank $n - 1$, $\boldsymbol{\Sigma}_{y^*}$ has rank $n - 1$. As a result, Thurstone's model assigns zero probabilities to all the $2^j - n!$ intransitive paired comparisons patterns (those patterns that do not correspond to ranking patterns). Thus, Thurstone's model is not a plausible model for multiple judgment paired comparisons data, but it may be a suitable model for ranking data (Maydeu-Olivares, 1999).

3. Thurstonian models appropriate for multiple judgment paired comparisons

We now describe two different solutions to the problem of specifying a Thurstonian model that assigns non-zero probabilities to all 2^j binary patterns.

3.1 The Thurstone-Takane model

Takane (1987) proposed adding a random error to each paired comparison so that $\mathbf{y}^* = \mathbf{A} \mathbf{t} + \mathbf{e}$. Furthermore, he assumed that

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{e} \end{pmatrix} \sim N \left(\begin{pmatrix} \boldsymbol{\mu}_t \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_t & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}^2 \end{pmatrix} \right) \quad (13)$$

where $\boldsymbol{\Omega}^2$ is a diagonal matrix with elements $\omega_{i,j}^2$.

Under this model, the probability of any paired comparisons pattern is given also by (6) and (8) but now $\boldsymbol{\mu}_{y^*} = \mathbf{A} \boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_{y^*} = \mathbf{A} \boldsymbol{\Sigma}_t \mathbf{A}' + \boldsymbol{\Omega}^2$. Furthermore, performing the change of variable of integration (9), we can write these pattern probabilities as (11) and (12) where instead of (10) we have

$$\boldsymbol{\tau} = -\mathbf{D} \mathbf{A} \boldsymbol{\mu}_t, \quad \mathbf{P}_{z^*} = \mathbf{D} (\mathbf{A} \boldsymbol{\Sigma}_t \mathbf{A}' + \boldsymbol{\Omega}^2) \mathbf{D} \quad (14)$$

Takane (1987) also proposed an interesting special case of this model in which it is assumed that $\boldsymbol{\Omega}^2 = \omega^2 \mathbf{I}$.

3.2 Thurstonian correlation structure models

The Thurstone and Thurstone-Takane models are mean and covariance structure models in the sense that they impose constraints on the mean vector and covariance matrix of \mathbf{y}^* . However, as only the binary choices are observed, the covariance matrix of \mathbf{y}^* can only be estimated from a correlation matrix. Moreover, to estimate the parameters of these models one must resort to pre and post-multiply the covariance structure by the inverse of a diagonal matrix of model-based standard deviations. This results in complex non-linear restrictions on the thresholds and tetrachoric correlations.

An alternative solution to the problem of specifying a proper model for multiple judgment paired comparison data within a Thurstonian framework is to specify restrictions on the means and correlations of \mathbf{y}^* while leaving the structure for the variances of \mathbf{y}^* unspecified. Now, the mean and covariance structures for \mathbf{y}^* implied by Thurstone's model are $\boldsymbol{\mu}_{\mathbf{y}^*} = \mathbf{A}\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_{\mathbf{y}^*} = \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}'$. Thus, we shall assume $\boldsymbol{\mu}_{\mathbf{y}^*} = \mathbf{A}\boldsymbol{\mu}_t$. Also, by parsimony we shall assume $\mathbf{P}_{\mathbf{y}^*} = \text{Off}(\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}')$, where $\text{Off}(\bullet)$ denotes the restrictions imposed on the off-diagonal elements of a correlation matrix. That is, we assume that the restrictions on the correlations among \mathbf{y}^* have the same functional form as the restrictions of Thurstone's model on the covariances among \mathbf{y}^* . Then, the probability of any paired comparisons pattern under this Thurstonian mean and correlation structure model is

$$\Pr\left(\bigcap_{l=1}^{\tilde{n}} y_l\right) = \int_{\mathbf{R}} \cdots \int_{\mathbf{R}} \phi_{\tilde{n}}(\mathbf{y}^* : \boldsymbol{\mu}_{\mathbf{y}^*}, \mathbf{P}_{\mathbf{y}^*}) d\mathbf{y}^* \quad (15)$$

with limits of integration (8). These pattern probabilities are unchanged when we transform \mathbf{y}^* using $\mathbf{z}^* = \mathbf{y}^* - \boldsymbol{\mu}_{\mathbf{y}^*}$. As a result, $\boldsymbol{\mu}_{\mathbf{z}^*} = \mathbf{0}$ and

$$\boldsymbol{\tau} = -\mathbf{A}\boldsymbol{\mu}_t \quad \mathbf{P}_{\mathbf{z}^*} = \text{Off}(\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}') \quad (16)$$

where now the pattern probabilities are given by (11), (12) and (16).

This model is more restrictive than the Thurstone-Takane model. However, unlike (14), the restrictions in (16) are linear.

4. Limited information estimation and testing

We shall now discuss minimal identification restrictions for the models just described. These minimal identification restrictions yield unrestricted models.

Given the comparative nature of the data, in all cases it is necessary to set the location of the elements of $\boldsymbol{\mu}_t$ and the location of the elements in each of the rows (columns) of $\boldsymbol{\Sigma}_t$. Arbitrarily, we set $\mu_{t1} = 0$ and $\boldsymbol{\Sigma}_t = \mathbf{P}_t$ (a matrix with ones along its diagonal). In addition, in the Thurstonian ranking model it is necessary to fix

one of the elements of \mathbf{P}_t (Dansie, 1986). Arbitrarily, we set $\rho_{n,n-1} = 0$. Also, in the Thurstone-Takane model, if $\mathbf{\Omega}^2$ is assumed to be diagonal it is necessary to set the location for its elements; arbitrarily we set $\omega_n^2 = 1$. Alternatively, if $\mathbf{\Omega}^2 = \omega^2 \mathbf{I}$ is assumed, we impose the constraint $\omega^2 = 1$.

Now, we collect the observed paired comparisons patterns in a \tilde{n} -way binary contingency table. We have seen that the models under consideration assume that the contingency table arises by dichotomizing a \tilde{n} -dimensional multivariate standard normal density according to a set of thresholds while imposing constraints -given by Eq. (10), (14), and (16)- on the thresholds and tetrachoric correlations. Muthén (1978, 1993) has proposed well established methods for estimating dichotomized multivariate normal structural models from the univariate and bivariate marginals of the contingency table. Here, we investigate whether Muthén's (1993) approach is suitable for these type of models by means of a simulation study.

In this estimation procedure, first each threshold is estimated separately from the univariate marginals of the contingency table via. Then, each of the tetrachoric correlations of the contingency table is estimated separately from the bivariate marginals of the contingency table given the estimated thresholds. Finally, the model parameters are estimated from the estimated thresholds and tetrachoric correlations using an unweighted least squares (ULS) function. His method yields asymptotically correct standard errors and tests of the structural restrictions imposed by the model on the thresholds and tetrachoric correlations. The latter are obtained by adjusting the minimum of the ULS fitting function by its mean (T_s) or by its mean and variance (T_a). Recently, Maydeu-Olivares (2001) has proposed a similar adjustment to the sum of squared residuals to the univariate and bivariate margins of the table, which gives us a limited information test of the overall restrictions imposed by the model on the contingency table. The degrees of freedom available for testing when fitting paired comparisons models is

$r = \frac{\tilde{n}(\tilde{n}+1)}{2} - q$, where q is the number of model parameters. When fitting ranking data it is necessary to adjust for the number of redundancies among the univariate and bivariate marginals which arise from having $2^{\tilde{n}} - n!$ structural empty cells in the contingency table. The number of degrees of freedom in this case is $r = \frac{\tilde{n}(\tilde{n}+1)}{2} - \sum_{x=2}^{n-1} \binom{x}{2} - q$ (Maydeu-Olivares, 1999).

We provide in Table 1 the simulation results obtained in estimating a model for 7 stimuli. Hence $\tilde{n} = 21$ binary variables were modeled. Sample size was 500. The true parameters were

$$\boldsymbol{\mu}_t = (0.5, 0, -0.5, 0, 0.5, -0.5, 0)' , \boldsymbol{\rho}_t = (0.8, 0.7, 0.6, 0.8, 0.7, 0.6, \dots, 0.8, 0.7, 0.6)'$$

where \mathbf{p}_t denotes the elements below the diagonal in \mathbf{P}_t . In the Thurstone-Takane model where $\mathbf{\Omega}^2$ was assumed to be diagonal, we used as true values $\mathbf{\Omega}^2 = \mathbf{I}$. 1000 replications were used for each of the three models estimated.

In this table we point out the largest relative bias (in %) observed in the parameter estimates and standard errors. We also provide the empirical rejection rates for the test statistics at $\alpha = 0.05$. They should be as close as possible to 0.05. Finally, we also provide the value of D_{KS} , the Kolmogorov-Smirnov one-sample test. The null hypothesis of a chi-square distribution for the empirical distribution of the test statistic is rejected at $\alpha = 0.05$ if $D_{KS} > 1.35$.

As can be seen in this table, for this particular choice of true generating values and sample size:

1. The performance of the tests statistics considered depends on the model. In all cases the mean and variance adjusted test statistics of overall restrictions matched well its reference chi-square distribution as indicated by D_{KS} .
2. Accurate parameter estimates and standard errors (largest relative bias $< |10\%$) were obtained for all models but for the Thurstone-Takane model when $\mathbf{\Omega}^2$ is assumed to be diagonal.

Table 1. Simulation results

model	maximum relative bias (%)						D_{KS} statistic (RR at $\alpha = 0.05$)			
	par. estimates			standard errors			structur. tests		overall tests	
	μ_t	\mathbf{P}_t	$\mathbf{\Omega}^2$	μ_t	\mathbf{P}_t	$\mathbf{\Omega}^2$	T_s	T_a	T_s	T_a
(A) Thurstone's	<1	<1	-	-3	-4	-	3.9 (13.9)	2.3 (4.2)	7.9 (26.3)	1.2 (6.4)
(B) TT, $\mathbf{\Omega}^2$ diagonal	2	-5	29	-8	-17	22	1.0 (5.5)	1.8 (2.2)	7.8 (23.0)	0.8 (5.6)
(C) TT, $\mathbf{\Omega}^2 = \omega^2 \mathbf{I}$	1	-1	-	-5	4	-	1.2 (6.1)	2.1 (2.4)	8.5 (25.2)	0.9 (5.0)
(D) correlation struct.	<1	<1	-	-1	-6	-	2.2 (9.4)	2.2 (3.8)	7.8 (24.6)	1.1 (4.3)

In Thurstone's model we fixed $\rho_{7,6} = .6$ (the true value) to calculate relative bias. Thurstone's model was fitted to ranking data, the other three models to paired comparisons. TT, Thurstone-Takane model; RR, rejection rate in %, should be close to 0.05.

Additional simulation studies with these true generating values reveal that accurate parameter estimates, standard errors and goodness of fit tests can be obtained for model (B) with 1000 observations, with model (A) with 300 observations, and with models (C) and (D) with as few as 100 observations.

5. Discussion and conclusions

Classical estimation procedures for Thurstonian models (e.g., Bock & Jones, 1968) estimate the model parameters using only the univariate means of the binary variables. These procedures are clearly unsatisfactory because:

1. The sample means are not independent. Standard errors and tests of the goodness of fit of the model computed under an independence assumption will be incorrect.
2. Unrestricted Thurstonian models and many restricted models are not identified from univariate information only. However, any Thurstonian model can be identified as soon as bivariate information is employed.

We have shown that using univariate and bivariate information it is possible to obtain accurate parameter estimates, standard errors and tests of goodness of fit to Thurstonian models for paired comparisons and ranking data with very small samples, even in large models. We have also pointed out that Thurstone's original model is suitable for ranking data, but that it is not a proper model for multiple judgment paired comparisons data. We have discussed two classes of Thurstonian models suitable for these data: a correlation structure model, and a covariance structure model (the Thurstone-Takane model). These two models are not equivalent because the Thurstone-Takane model is not scale invariant. Although the choice between a Thurstonian covariance vs. a correlation structure model should be substantively motivated, it is also important to consider model fit and estimation issues in choosing between these models.

References

- Bock, R.D. & Jones, L.V. (1968). *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day.
- Dansie, B.R. (1986) Normal order statistics as permutation probability models. *Applied Statistics*, 35, 269-275.
- Maydeu-Olivares, A. (1999). Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika*, 64, 325-340.
- Maydeu-Olivares, A. (2001). Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgment sampling. *Psychometrika*, 66, 209-228.
- 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.
- Takane, Y. (1987). Analysis of covariance structures and probabilistic binary choice data. *Communication and Cognition*, 20, 45-62.
- Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review*, 79, 281-299.
- Thurstone, L.L. (1931). Rank order as a psychological method. *Journal of Experimental Psychology*, 14, 187-201.