Chapter 3

Thurstone's Case V model: A structural equations modeling perspective

Albert Maydeu-Olivares

University of Barcelona, Faculty of Psychology.
Barcelona, Spain.
Instituto de Empresa, Marketing Department.
Madrid, Spain.

3.1 Introduction

Modeling how we choose among alternatives, or more generally, modeling preferences, is one of the core topics of study in Psychology. Preferences can be studied experimentally using a variety of procedures, one of the oldest being the method of paired comparisons. This method remains quite popular in areas such as psychophysics and consumer psychology. For a good overview of the method of paired comparisons see David (1988).

The most common paired comparisons procedure is what Bock and Jones (1968) called multiple judgment paired comparisons. Suppose preferences for n stimuli are to be investigated. In this procedure we collect a random sample of individuals from the population we wish to investigate, we construct all possible paired comparisons, and all individuals in the sample are presented all pairs one at a time. For each pair, each individual is asked to choose one stimuli within each pair and his or her response is coded using a binary variable. Since there are $\tilde{n} = \binom{n}{2} = \frac{n(n-1)}{2}$ paired comparisons with n objects, for each individual we obtain a pattern of $\tilde{n}$ binary

K.van Montfort et al. (eds), Recent Developments on Structural Equation Models, 41-67.

observations. The objective is then to model the set of possible $2^{\tilde{n}}$ paired comparison patterns. Some of the these patterns may be intransitive, while others are transitive. A pattern of paired comparisons is said to be transitive when given the pattern it is possible to order the individual preferences, and intransitive otherwise. For example, consider three stimuli, $\{i, j, k\}$, and suppose an individual chooses *i* over *j*, *i* over *k*, and *j* over *k*. This pattern of binary choices is transitive as the stimulus most preferred by this individual is *i*, the second most preferred stimulus is *j*, and the least preferred stimulus is *k*. Suppose, on the other hand, that the individual chooses *i* over *j*, *i* over *k*, and *k* over *j*. This pattern is intransitive as we can not order the preferences of this individual for these stimuli given these binary choices. Since the total number of orderings of *n* stimuli is *n*!, this is number of possible transitive patterns in a multiple judgment paired comparisons experiment. The number of possible intransitive paired comparisons patterns is obviously $2^{\tilde{n}} - n!$.

An alternative method to study preferences is the ranking method, which is also quite popular in applications. In this method, all stimuli are presented at once to the respondents, and the respondents are asked to rank or to order the stimuli according to their preferences. The multiple judgment paired comparisons method and the ranking method are closely related. This is because we can transform the observed rankings to patterns of paired comparisons. However, since the paired comparisons patterns arising from a ranking experiment must be transitive, only *n*! paired comparisons patterns can be observed.

One of the oldest models for paired comparisons data is Thurstone's (1927) law of comparative judgment. Arguably it remains the most influential model to date along with Luce's (1959) choice model (see Böckenholt, 2001). In a nutshell, Thurstone assumed that whenever two stimuli are presented to an individual, each stimulus elicits an unobserved continuous preference (*discriminal process* in Thurstone's terminology) which is normally distributed, and that the individual chooses the stimuli with the largest continuous preference. To model ranking data, Thurstone (1931) proposed transforming the observed ranking patterns to patterns of binary paired comparisons and fitting his paired comparisons model (Thurstone, 1927) to the transformed data. In his 1927 seminal article, Thurstone described in detail a variety of special cases and restricted versions of his model. Perhaps the most popular restricted version of Thurstone's model is his Case V model. In this model, Thurstone assumed that the continuous preferences were uncorrelated and had common variance. In recent years, more complex restricted Thurstonian models have been proposed. For a good overview of restricted Thurstonian models, see Takane (1987).

It turns out that Thurstone's model (and in particular, Thurstone's Case V model) is not a proper model for multiple judgment paired comparisons data (Maydeu-Olivares, 1999). This is because under Thurstone's model intransitive patterns have zero probability. In other words, according to Thurstone's model all respondents must be transitive. This is obviously an implausible assumption for multiple judgment data. Thurstone's model, on the other hand is a plausible model for paired comparisons obtained via a ranking experiment, as in this case the respondents are forced to be transitive by the use of the ranking method. It was not until 1987 that Takane proposed an extension of Thurstone's model for paired comparisons that accounts for the intransitive patterns that may be observed in a multiple judgment paired comparisons experiment. In this paper we shall refer to Takane's (1987) extension of Thurstone's (1927) model as the Thurstone-Takane model.

Despite their theoretical appeal, estimating Thurstonian models for multiple judgment paired comparisons and ranking data is in principle involved as to compute a pattern probability under these models it is necessary to integrate a high dimensional normal density. Because it is difficult to evaluate these integrals, theoretical research on Thurstonian modeling of multiple judgment paired comparisons and of ranking data has stagnated for years. Also, by and large applied researchers seem to continue using the simplest and most restrictive versions of Thurstone's model (such as his Case V) as well as the simplest estimation approaches to these models, such as the classical approach described in Torgerson (1958). Recently, with the development of new statistical methods for handling multivariate normal integrals and the advent of fast computers we see a renewed interest in these models and in these data (Böckenholt, 1990, 1993; Brady, 1989; Chan & Bentler, 1998; Maydeu-Olivares, 1999, 2001; Tsai & Böckenholt, 2001; Tsai & Yao, 2000; Yao & Böckenholt, 1999; Yu, 2000).

In this paper we show that estimating Thurstone's Case V model (suitable for ranking data) and estimating the Thurstone-Takane Case V model (suitable for multiple judgment paired comparisons data) is similar to the problem of estimating a factor model from binary data. This model assumes that a multivariate normal density with a factor structure has been dichotomized according to a set of thresholds. Thus, to compute a pattern probability under this model it is also necessary to integrate a high dimensional normal density. However, the factor model for binary data can be straightforwardly estimated using software for structural equation modeling with capabilities for handling binary data such as MPLUS (Muthén & Muthén, 1998) without integrating high dimensional normal densities. Rather, within a structural equations framework the factor model is estimated as follows: First, the thresholds and tetrachoric correlations are estimated. Then, if no restrictions are imposed on the thresholds, the factor loadings are estimated from the tetrachoric correlations. Alternatively, if some structure is assumed on the thresholds, then the model parameters are estimated in the second stage from the thresholds and tetrachoric correlations. The purpose of this paper is to show that this structural equations approach can also be applied to estimate Thurstonian Case V models to ranking and to paired comparisons data. In fact, these models can be as straightforwardly estimated as a factor model for binary data. Thus, applied researchers can use widely available structural equations modeling software to draw sound statistical inferences from paired comparisons and ranking data.

The remaining of this article is structured in three sections. In the next section Thurstone's Case V model and the Thurstone-Takane model are presented and we provide the restrictions imposed by these models on the thresholds and tetrachoric correlations. In the third section we provide the relationship between these models and the factor model for binary data. In this section we also describe how to estimate Thurstone's Case V model for ranking data and the Thurstone-Takane model for paired comparisons data using MPLUS. Two examples are provided. In the first example we model purchasing preferences for compact cars collected using paired comparisons. In the second example we model career preferences among Psychology undergraduates collected using rankings.

An added benefit of employing an structural equations approach to model paired comparisons and ranking data is that one can incorporate to the model background information on the respondents. This is the topic of the fourth section of the manuscript. In this section we re-estimate the compact cars' paired comparisons data using MPLUS including in the model the gender, age, and family income of the respondents.

As an appendix we describe the classical estimation procedure for Thurstone's Case V model (Mosteller, 1951a; Torgerson, 1958) that may be familiar to some readers, we discuss its limitations, and we relate it to the structural equations approach employed here.

3.2 Thurstone's Case V model

In this section, we start by presenting Thurstone's Case V model for one paired comparison as is generally presented in the literature.

3.2.1 Case V model Thurstone's for one paired comparison

Suppose we wish to investigate how the members of a population choose between two stimuli, i and j. We collect a random sample of N individuals from that population and we present each individual both stimuli asking him or her to choose one stimuli. The individuals' responses are then coded as follows:

$$y_{i,j} = \begin{cases} 1 & \text{if stimulus } i \text{ is chosen} \\ 0 & \text{if stimulus } j \text{ is chosen} \end{cases} \tag{1}$$

Thus, we obtain a binary variable and we wish to model $\Pr(y_{i,j} = 1)$ and $\Pr(y_{i,j} = 0)$. To model these probabilities, Thurstone's (1927) law of comparative judgment introduces the following assumptions:

(a) Each respondent has a continuous preference t_i for stimulus *i* and a continuous preference t_j for stimulus *j*.
(b) Both continuous preferences t_i and t_j are normally distributed in the population.
(c) A respondent will choose stimulus *i* if his/her continuous preference for this stimulus is greater that his/her continuous preference for stimulus *j*, otherwise s/he will choose stimulus *j*.

Thurstone's Case V is a special case of this general model in which it is further assumed that

(d) The continuous preferences t_i and t_j are uncorrelated in the population and they have a common variance σ^2.

Thustone's Case V model implies that

$$\Pr(y_{i,j} = 1) = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{2\sigma^2}}\right) \tag{2}$$

where $\Phi(\bullet)$ denotes a univariate standard normal distribution function, and μ_i and μ_j denote the mean of the continuous preferences for stimuli *i* and *j* in the population of interest. Obviously, $\Pr(y_{i,j} = 0) = 1 - \Pr(y_{i,j} = 1)$.

We shall now present how one reaches (2) from assumptions (a) to (d) to better understand the case in which more than one paired comparison is modeled. To do so, we write $\mathbf{t} = (t_i, t_j)'$. Then, from assumptions (b) and (d)

$$\mathbf{t} \sim N\left(\boldsymbol{\mu}, \sigma^2\mathbf{I}\right) \tag{3}$$

Now, following Thurstone (1927) we take the difference between the unobserved preferences

$$y^*_{i,j} = t_i - t_j \,. \tag{4}$$

Then, assumption (c) implies that

$$y_{i,j} = \begin{cases} 1 & \text{if} \quad y^*_{i,j} \geq 0 \\ 0 & \text{if} \quad y^*_{i,j} < 0 \end{cases} \tag{5}$$

Finally, equations (3), (4) and (5) imply that under Thurstone's Case V model

$$\Pr(y_{i,j} = 1) = \int_0^\infty \phi_1(y^*_{i,j} : \mu_i - \mu_j, 2\sigma^2) dy^*_{i,j} \,, \tag{6}$$

and $\Pr(y_{i,j} = 0) = \int_{-\infty}^0 \phi_1(y^*_{i,j} : \mu_i - \mu_j, 2\sigma^2) dy^*_{i,j}$, where $\phi_n(\bullet)$ denotes a n-variate normal density function.

Equation (2) is obtained from (6) by standardizing $y^*_{i,j}$. This leaves the probabilities unchanged. Let

$$z^*_{i,j} = \frac{y^*_{i,j} - \mu_{y^*_{i,j}}}{\sigma_{y^*_{i,j}}} \,, \tag{7}$$

where $\mu_{y^*_{i,j}}$ and $\sigma_{y^*_{i,j}}$ denote the mean and standard deviation of $y^*_{i,j}$. Then, when $y^*_{i,j} = 0$, $z^*_{i,j}$ takes the value

$$\frac{0 - \mu_{y^*_{i,j}}}{\sigma_{y^*_{i,j}}} = \frac{-(\mu_i - \mu_j)}{\sqrt{2\sigma^2}} = \tau_{i,j} \,, \tag{8}$$

which we denote by $\tau_{i,j}$. Also, the mean and variance of $z^*_{i,j}$ are 0 and 1, respectively. Equation (2) then follows immediately:

$$\Pr(y_{i,j} = 1) = \int_{\tau_{i,j}}^\infty \phi_1(z^*_{i,j} : 0,1) dz^*_{i,j} = \Phi(-\tau_{i,j}) = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{2\sigma^2}}\right). \tag{9}$$

We now turn to the case where we are interested in modeling preferences for $n > 2$ stimuli using a paired comparisons design. In this case we are to model the probability of observing a pattern of paired comparisons. This probability is obtained by integrating a multivariate normal density. We shall see what restrictions Thurstone's Case V model imposes on the thresholds and tetrachoric correlations of a multivariate normal density. Later on, we shall see that these restrictions are very similar to those imposed by a factor model. This similarity makes straightforward to estimate Thurstone's Case V model within a structural equations approach.

3.2.2 Thurstone's Case V model for multiple paired comparisons

When preferences for n stimuli are to be modeled there are $\tilde{n}$ paired comparisons. To investigate preferences for these stimuli in a population most often we collect a random sample of respondents and we present each respondent all pairs, one pair at a time, asking the respondents to choose one stimulus within each pair. To avoid order effects, the experimenter must randomize the order of presentation of the pairs, as well as the order of stimuli within each pair. The paired comparisons obtained by this procedure have been termed *multiple judgement paired comparisons* by Bock and Jones (1968).

Now, using (1), for each respondent we obtain a pattern of $\tilde{n}$ binary observations. The objective now is to model the probability of observing each of the possible $2^{\tilde{n}}$ binary patterns. To express the pattern probabilities under Thurstone's Case V model it is convenient to use matrix notation.

We write (4) in matrix notation as

$$\mathbf{y}^* = \mathbf{At} \tag{10}$$

where $\mathbf{t}$ is a $n \times 1$ vector given by (3), $\mathbf{y}^*$ is a $\tilde{n} \times 1$ vector, and $\mathbf{A}$ 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. When n = 2, $\mathbf{A} = (1\ \ -1)$, whereas when n = 3 and n = 4,

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \tag{11}$$

respectively. Thus, the first row of $\mathbf{A}$ correspond to the comparison between stimulus one and two, the second row between stimulus one and three, and so forth.

The probability of observing any pattern of paired comparisons under Thurstone's Case V model is obtained by putting together (3), (10) and (5). This probability is

$$\Pr\left(\bigcap_{i,j} y_{i,j}\right) = \int \underset{\mathbf{R}}{\cdots} \int \phi_{\tilde{n}}\left(\mathbf{y}^* : \boldsymbol{\mu}_{y^*}, \boldsymbol{\Sigma}_{y^*}\right) d\mathbf{y}^* \tag{12}$$

where the limits of integration are $(0,\infty)$ if $y_{i,j} = 1$ and $(-\infty,0)$ if $y_{i,j} = 0$. The mean vector and covariance matrix of $\mathbf{y}^*$ are readily obtained from (10) and (3)

$$\boldsymbol{\mu}_{y^*} = \mathbf{A}\boldsymbol{\mu} \qquad \boldsymbol{\Sigma}_{y^*} = \sigma^2 \mathbf{A}\mathbf{A}'. \tag{13}$$

We shall provide an example to better understand Equation (12). Suppose that we are interested in modeling three stimuli, $\{i, j, k\}$. Then all possible paired comparisons are $\{i, j\}$, $\{i, k\}$ and $\{j, k\}$. The probability that an individual prefers i over j, i over k, but k over j is under Thurstone's Case V model,

$$\Pr\left[\left(y_{i,j} = 1\right) \cap \left(y_{i,k} = 1\right) \cap \left(y_{j,k} = 0\right)\right] = \int_0^\infty \int_0^\infty \int_{-\infty}^0 \phi_3\left(\mathbf{y}^* : \boldsymbol{\mu}_{y^*}, \boldsymbol{\Sigma}_{y^*}\right) d\mathbf{y}^* \tag{14}$$

where by (13),

$$\boldsymbol{\mu}_{y^*} = \begin{pmatrix} \mu_i - \mu_j \\ \mu_i - \mu_k \\ \mu_j - \mu_k \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{y^*} = \begin{pmatrix} 2\sigma^2 & & \\ \sigma^2 & 2\sigma^2 & \\ -\sigma^2 & \sigma^2 & 2\sigma^2 \end{pmatrix}. \tag{15}$$

As it can be seen in this last equation, the binary choice probabilities $\Pr(y_{i,j} = 1)$, $\Pr(y_{i,k} = 1)$ and $\Pr(y_{j,k} = 1)$ are not independent under Thurstone's Case V model. This is important, as classical estimation procedures for Thurstonian models -see Appendix- assume that binary choice probabilities are independent.

When there were only two stimuli we standardized $\mathbf{y}^*$ using (7) transforming (6) to obtain (2). We shall now standardize $\mathbf{y}^*$ when $n > 2$. In matrix notation, (7) can be written as

$$\mathbf{z}^* = \mathbf{D}\left(\mathbf{y}^* - \boldsymbol{\mu}_{y^*}\right), \qquad \mathbf{D} = \left(\mathrm{Diag}\left(\boldsymbol{\Sigma}_{y^*}\right)\right)^{-\frac{1}{2}}. \tag{16}$$

That is, $\mathbf{D} = \left(\text{Diag}\left(\sigma^2\mathbf{AA}'\right)\right)^{-\frac{1}{2}}$. As a result $\boldsymbol{\mu}_{y^*} = \mathbf{0}$ and

$$\mathbf{P}_{z^*} = \mathbf{D\Sigma}_{y^*}\mathbf{D} = \mathbf{D}\left(\sigma^2\mathbf{AA}'\right)\mathbf{D} = \frac{1}{2}\mathbf{AA}'. \tag{17}$$

After standardization, the pattern probabilities (12) can be equivalently written as

$$\Pr\left(\bigcap_{i,j} y_{i,j}\right) = \int \underset{\mathbf{R}}{\cdots} \int \phi_{\tilde{n}}\left(\mathbf{z}^* : \mathbf{0}, \mathbf{P}_{z^*}\right) d\mathbf{z}^* \tag{18}$$

where the limits of integration are now $(\tau_{i,j}, \infty)$ if $y_{i,j} = 1$ and $(-\infty, \tau_{i,j})$ if $y_{i,j} = 0$. The thresholds $\tau_{i,j}$ were defined in (8). If we stack all thresholds in a vector, the restrictions imposed by Thurstone's Case V model on the thresholds can be written as

$$\boldsymbol{\tau} = -\mathbf{D}\boldsymbol{\mu}_{y^*} = -\mathbf{DA}\boldsymbol{\mu}. \tag{19}$$

Also we notice that since $\mathbf{z}^*$ is a multivariate normal density and (17) has ones along its diagonal, the elements of this matrix are tetrachoric correlations. Following our previous example, the restrictions imposed by Thurstone's Case V model on the thresholds (19) and tetrachoric correlations (17) for three stimuli are

$$\boldsymbol{\tau} = \begin{pmatrix} \dfrac{-\mu_i + \mu_j}{\sqrt{2\sigma^2}} \\ \dfrac{-\mu_i + \mu_k}{\sqrt{2\sigma^2}} \\ \dfrac{-\mu_j + \mu_k}{\sqrt{2\sigma^2}} \end{pmatrix} \qquad \mathbf{P}_{z^*} = \begin{pmatrix} 1 & & \\ 1/2 & 1 & \\ -1/2 & 1/2 & 1 \end{pmatrix}. \tag{20}$$

It is interesting to notice that under Thurstone's Case V model the tetrachoric correlations are patterned, but that they do not depend on any model parameters.

Yet, Thurstone's model for paired comparisons data assigns zero probability to all intransitive patterns (Maydeu-Olivares, 1999). Thus, Thurstone's model is not a proper model for multiple judgment paired comparisons data and should not be used to model these kind of data. On the other hand, precisely because it assigns zero probabilities to all intransitive patterns it is a proper model for paired comparisons obtained from ranking data. We now discuss the application of Thurstone's Case V model to ranking data.

3.2.3 Thurstone's Case V model for ranking data

In a ranking experiment, all n stimuli are presented to the respondents at once. The respondents are asked to rank the stimuli according to their preferences. Thurstone (1931) proposed a model for ranking data that simply consists in transforming the observed rankings to paired comparisons and applying his model for paired comparisons. The rankings are transformed to paired comparisons by constructing a dichotomous variable $y_{i,j}$ for each ordered pairwise combination of stimuli to indicate which stimulus was ranked above the other

$$y_{i,j} = \begin{cases} 1 & \text{if stimulus } i \text{ is ranked above stimulus } j \\ 0 & \text{if stimulus } i \text{ is ranked below stimulus } j \end{cases} . \qquad (21)$$

Then, after transforming the observed rankings to binary data using (21) the probability of observing any of the $n!$ possible ranking patterns under Thurstone's Case V model is given by (18) with (19) and (17). Thurstone's original Case V model is a proper model for rankings as it assigns zero probabilities to all intransitive paired comparisons patterns. Intransitive patterns can not be observed because in a ranking experiment respondents are "forced" to be transitive.

The fact that when the binary patterns are obtained from ranking patterns only $n!$ patterns can be observed instead of $2^{\tilde{n}}$ introduces

$$r = \sum_{x=2}^{n-1} \binom{x}{2} \qquad (22)$$

redundancies among the thresholds and tetrachoric correlations estimated from the binary variables (Maydeu-Olivares, 1999). For this reason, the number of degrees of freedom when estimating a Thurstonian model from ranking data does not equal the number of thresholds plus the number of tetrachoric correlations minus the number of estimated parameters, i.e. $c = \tilde{n} + \tilde{n}(\tilde{n}-1)/2 - q$. Rather the correct number of degrees of freedom is $df = c - r$. For convenience, we list r as a function of n in a table for n between 3 and 10.

n	3	4	5	6	7	8	9	10
r	1	4	10	20	35	56	84	120

We now present an extension of Thurstone's model due to Takane (1987) aimed at accounting for the intransitivities that may be observed in multiple judgment paired comparisons experiments.

3.2.4 Thurstone-Takane model for multiple judgement paired comparisons data

In the Thurstone-Takane model, an error is added to each paired comparison reflecting that a respondent's preference for a stimulus can change during the paired comparisons experiment as the stimulus is presented next to different stimuli, thus inducing the intransitivities. In other words, while in Thurstone's original model (10) is assumed, in the Thurstone-Takane model it is assumed instead that

$$\mathbf{y}^* = \mathbf{A}\mathbf{t} + \mathbf{e}\,. \tag{23}$$

Furthermore, in the Thurstone-Takane model it is assumed that the paired specific errors $\mathbf{e}$ are normally distributed with zero means and common variance ω^2, and that they are mutually uncorrelated and uncorrelated with the preferences $\mathbf{t}$. In other words, while in Thurstone's original Case V model (3) is assumed, in the Thurstone-Takane Case V model it is assumed instead that

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{e} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_t \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \omega^2\mathbf{I} \end{pmatrix} \right). \tag{24}$$

Now, putting together (23), (24), and (5) the probability of any given paired comparisons pattern under the Thurstone-Takane Case V model is given by (12), where instead of (13), the mean vector and covariance matrix of $\mathbf{y}^*$ are $\boldsymbol{\mu}_{y^*} = \mathbf{A}\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}_{y^*} = \sigma^2\mathbf{A}\mathbf{A}' + \omega^2\mathbf{I}$, respectively. Also, as we did with Thurstone's original Case V model, we can apply the standardization (16) to the Thurstone-Takane Case V model so as to express the pattern probabilities under this model using (18) where now the restrictions imposed on the thresholds and tetrachoric correlations are

$$\boldsymbol{\tau} = -\mathbf{D}\mathbf{A}\boldsymbol{\mu} \tag{25}$$

$$\mathbf{P}_{z^*} = \mathbf{D}\left(\sigma^2\mathbf{A}\mathbf{A}' + \omega^2\mathbf{I}\right)\mathbf{D} \tag{26}$$

with $\mathbf{D} = \left(\mathrm{Diag}\left(\sigma^2 \mathbf{AA}' + \omega^2 \mathbf{I}\right)\right)^{-\frac{1}{2}}$. Under this model, all paired comparisons patterns have non-zero probability and thus it is a suitable model for multiple judgment paired comparisons data.

To sum up, we have described two models, Thurstone's Case V model and the Thurstone-Takane Case V model, and we have provided the restrictions that these models impose on the thresholds and tetrachoric correlations. The former is suitable for modeling ranking data, whereas the latter is suitable for modeling multiple judgment paired comparisons data. In the next section we describe how these models can be estimated within a structural equations framework and we provide two numerical examples.

3.3 Structural Equation Modeling Of Thurstone's Case V Model

There is a close correspondence between the models discussed here and the factor model for binary data. In fact, estimating these models within a structural equations framework is as straightforward as estimating a factor model for binary data. Before we present the relationships between these models and the factor model it is convenient to provide identification restrictions for the models under consideration.

3.3.1 Identification restrictions for Thurstonian Case V models for paired comparisons and ranking data

Thurstone's Case V model imposes the restrictions (19) on the thresholds of the binary variables, and it imposes the restrictions (17) on the tetrachoric correlations. Not all parameters of the model are identified. Because of the comparative nature of the data, one of the means must be fixed to set the scale for the remaining means. To identify the model we shall set $\mu_n = 0$. Also, in this model σ^2 is not identified. It is convenient to set $\sigma^2 = \frac{1}{2}$ as when σ^2 is fixed at this value (19) and (17) become

$$\tau = -\mathbf{A}\mu \qquad \mathbf{P}_{z^*} = \frac{1}{2}\mathbf{AA}' \tag{27}$$

Similarly, the Thurstone-Takane Case V model imposes the restrictions (25) on the thresholds of the binary variables, and it imposes the restrictions (26) on the tetrachoric correlations. Not all parameters of this model are identified, either. Again, one of the means must be fixed to set the scale for the remaining means. To identify the model we shall set $\mu_n = 0$. Also, the variance of the paired specific errors is not identified. Hence we shall set $\omega^2 = 1$.

However, the Thurstone-Takane Case V model imposes non-linear restrictions on the thresholds and tetrachoric correlations through the model-based diagonal matrix **D**. These type of non-linear restrictions are not straightforward to implement in standard structural equations software. Fortunately, it is possible to obtain an equivalent model that does not involve **D** by a one-to-one reparamaterization of the Thurstone-Takane Case V model. Instead of using the set of identified parameters $\boldsymbol{\theta} = \{\mu_1, \cdots, \mu_{n-1}, \sigma^2\}$ we shall use the set of identified parameters $\tilde{\boldsymbol{\theta}} = \{\tilde{\mu}_1, \cdots, \tilde{\mu}_{n-1}, \tilde{\sigma}^2\}$ where

$$\tilde{\mu}_i = \frac{\mu_i}{\sqrt{1+2\sigma^2}} \qquad \tilde{\sigma}^2 = \frac{\sigma^2}{1+2\sigma^2}. \tag{28}$$

Applying this reparameterization to (25) and (26) we can write these equations as

$$\boldsymbol{\tau} = -\mathbf{A}\tilde{\boldsymbol{\mu}} \qquad \mathbf{P}_{z^*} = \tilde{\sigma}^2\mathbf{A}\mathbf{A}' + \mathbf{U} \tag{29}$$

where

$$\mathbf{U} = \mathbf{I} - \text{Diag}\left(\tilde{\sigma}^2\mathbf{A}\mathbf{A}'\right) = \left(1 - 2\tilde{\sigma}^2\right)\mathbf{I}. \tag{30}$$

This equivalent version of the Thurstone-Takane Case V model is straightforward to implement in standard computer programs for structural equation modeling.

3.3.2 Relationship between Thurstonian Case V models for paired comparisons and ranking data and the factor model for binary data

Suppose $\tilde{n}$ binary indicators are to be modeled using a p-factor model and the observed data is treated as categorical. Under this model (also known as multidimensional normal ogive model in the educational testing literature), the probability of observing a pattern of binary variables is also given by (18). In this model, the thresholds τ are left unconstrained and the following structure is assumed on the matrix of tetrachoric correlations, $\mathbf{P}_{z^*} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$, where Λ is a $\tilde{n} \times p$ matrix of factor loadings, $\boldsymbol{\Psi}$ is a $p \times p$ matrix of interfactor correlations, and Θ is a $\tilde{n}$ x $\tilde{n}$ diagonal matrix consisting of the variances of the unique factors where for identification purposes, $\boldsymbol{\Theta} = \mathbf{I} - \text{Diag}(\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}')$.

We immediately see some similarities between a factor model for $\tilde{n}$ binary variables and the Thurstonian models we are discussing. The similarities increase when we consider a factor model for binary data in which the factor means $\boldsymbol{\alpha}$ are free parameters to be estimated. Letting again $\boldsymbol{\Theta} = \mathbf{I} - \mathrm{Diag}(\Lambda\Psi\Lambda')$, the thresholds and tetrachoric correlations implied by a factor model with non-zero factor means are

$$\boldsymbol{\tau} = -\boldsymbol{\Lambda\alpha} \qquad\qquad \mathbf{P}_{z^*} = \Lambda\Psi\Lambda' + \Theta. \tag{31}$$

Comparing this equation to (27) we see that Thurstone's Case V model for *n* paired comparisons is equivalent to a factor model for *ñ* binary indicators with non-zero factor means if

(a) The number of factors is always *n*.
(b) The matrix of factor loadings equal the matrix of constants **A**.
(c) The factors are uncorrelated.
(d) The factors have common variance ½.
(e) There are no unique factors, so that **Θ** = **0**.

Also, comparing (31) to (29) we see that the Thurstone-Takane Case V model is even more similar to a factor model with non-zero factor means. In fact, a Thurstone-Takane Case V model is equivalent to a factor model for *ñ* binary indicators with non-zero factor means if

(a) The number of factors is always *n*.
(b) The factor loadings equal the matrix of constants **A**.
(c) The factors are uncorrelated.
(d) The factors have common variance σ^2.

3.3.3 Estimation of Thurstonian Case V models using MPLUS

To estimate these models in MPLUS, the observed binary variables are to be declared categorical. Also, we must let MPLUS know that we want to estimate a model with a mean/threshold structure. Estimation then proceed as follows. First, the thresholds and tetrachoric correlations are estimated. Then, collecting all the estimated thresholds and tetrachoric correlations in a vector $\boldsymbol{\kappa}$, and collecting all the *q* model parameters in a vector $\boldsymbol{\theta}$, the model parameters are estimated by minimizing

$$F = \left(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\boldsymbol{\theta})\right)' \hat{\mathbf{W}}\left(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\boldsymbol{\theta})\right). \tag{32}$$

Letting Ξ be the asymptotic covariance matrix of the sample thresholds and tetrachoric correlations, two obvious choices of $\hat{\mathbf{W}}$ in (32) are $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$ (WLS: Muthén, 1978, 1984), and $\hat{\mathbf{W}} = (\mathrm{Diag}(\hat{\Xi}))^{-1}$ (DWLS: Muthén, du Toit & Spisic, 1997). Asymptotically correct parameter estimates and standard errors can be obtained for both estimation methods. Also, the restrictions imposed by the model on the thresholds and tetrachoric correlations are tested as follows: For WLS estimation $N\hat{F}$ is employed which is asymptotically distributed as a chi-square distribution. In the case of DWLS estimation, a goodness of fit test can be obtained using either Satorra and Bentler's (1994) mean correction or their mean and variance correction to $N\hat{F}$ (see Muthén, 1993). As shown by Muthén (1993), larger samples are needed to obtain adequate parameter estimates, standard errors and goodness of fit tests with WLS than with DWLS. Thus, DWLS is generally to be recommended in practical applications.

Now, suppose we have performed a multiple judgment paired comparisons experiment with n stimuli. We wish to fit the Thurstone-Takane Case V model to the observed $\tilde{n}$ binary variables as this model assigns non-zero probabilities to all possible binary patterns (transitive and intransitive). To estimate this model with MPLUS we need only specify a factor model with n factors where

a) the correlations among these factors are zero,
b) the factor loadings are fixed constants given by the **A** matrix,
c) the factor variances are set equal,
d) the factor means are set free (except for the last factor mean, which is set to zero for identification purposes).

We need not be concerned with the diagonal elements of the matrix **U** in (29) as these elements are not included in the function to be minimized. MPLUS in addition can estimate of a set of intercepts in the mean structure. These must be set to zero.

Suppose that alternatively we have performed a ranking experiment with n stimuli. We then transform the rankings to $\tilde{n}$ binary variables using (21). We wish to fit Thurstone's Case V model to these binary variables as this model assigns non-zero probabilities only to transitive binary patterns. To estimate this model with MPLUS we need only specify a factor model with n factors with the same constraints we used for the Thurstone-Takane model except that the factor variances are now set equal to ½ for identification purposes. Also, when fitting ranking data the degrees of freedom provided by MPLUS are incorrect as MPLUS assumes that there are no redundancies among the sample thresholds and tetrachoric correlations. To test the fit of a ranking model estimated by WLS one needs to take $N\hat{F}$ and manually obtain a p-value using the number of degrees of freedom printed in the MPLUS output minus r -given in (22). The same procedure can be used when the model is estimated using DWLS and a mean scaled goodness of fit statistic. It is not possible to correct the mean and variance adjusted statistic in this way. Therefore, this statistic should not be employed when fitting ranking data using MPLUS.

We now provide two numerical examples to illustrate our presentation. In the first example the data are multiple judgment paired comparisons, while in the second example the data are rankings.

3.3.4 Modeling preferences for compact cars collected using paired comparisons

Maydeu-Olivares (2001) provided data on purchasing preferences for these four compact cars: {O = Opel Corsa, R = Renault Clio, S = Seat Ibiza, V = Volkswagen Polo}. Because there are six paired comparisons with four stimuli, six binary variables were observed for each respondent. These data will be re-analyzed here but only for the male respondents. This is because we found that preferences for men and women to these stimuli are rather different. There were 118 male respondents in the sample. Their responses to these four cars are provided in Table 1.

*Table 1.*Observed frequencies of paired comparisons patterns in the compact cars data

#	pattern	fq.	#	pattern	fq.	#	pattern	fq.	#	pattern	fq.
* 1	111111	7	17	101111	0	*33	011111	2	*49	001111	2
* 2	111110	9	18	101110	0	*34	011110	6	50	001110	1
3	111101	0	19	101101	0	35	011101	0	51	001101	0
* 4	111100	6	20	101100	0	36	011100	1	52	001100	0
* 5	111011	3	*21	101011	3	37	011011	0	*53	001011	1
6	111010	0	22	101010	0	38	011010	0	54	001010	1
* 7	111001	6	*23	101001	4	39	011001	0	55	001001	0
* 8	111000	7	24	101000	0	40	011000	0	56	001000	0
9	110111	0	25	100111	0	41	010111	1	*57	000111	3
10	110110	0	26	100110	0	*42	010110	2	*58	000110	1
11	110101	0	27	100101	0	43	010101	0	59	000101	0
*12	110100	17	28	100100	2	*44	010100	3	*60	000100	3
13	110011	0	29	100011	1	45	010011	0	*61	000011	1
14	110010	0	30	100010	1	46	010010	1	62	000010	1
15	110001	0	*31	100001	2	47	010001	0	*63	000001	0
*16	110000	12	*32	100000	5	48	010000	1	*64	000000	2

Notes: N = 118; * transitive patterns; the binary variables denote the responses to these paired comparisons {O,R}, {O,S}, {O,V}, {R,S}, {R,V}, and {S,V}, where O = Opel Corsa, R = Renault Clio, S = Seat Ibiza, V =Volkswagen Polo. The binary variables take a value of 1 if the first stimuli within a pair was chosen, and 0 otherwise.

As it can be seen in this table 11 respondents (roughly 9% of the sample) provided intransitive patterns. Thus, the vast majority of respondents in the sample are transitive. Yet, according to Thurstone's model all respondents in the population -not just in the sample- must be transitive.

We fitted a Thurstone-Takane Case V model (29) to these data using MPLUS. Since our sample is rather small we used DWLS estimation. Goodness of fit was assessed using the mean corrected Satorra-Bentler statistic, obtaining 15.87 on 17 degrees of freedom, p = 0.53 so this model fits the data rather well. $\tilde{\mu}_V$ was fixed at zero identification purposes. The remaining parameter estimates with standard errors in parentheses were $\tilde{\mu}_0$ = 0.09 (0.10), $\tilde{\mu}_R$ = -0.36 (0.11), $\tilde{\mu}_S$ = -0.50 (0.11), $\tilde{\sigma}^2$ = 0.46 (0.02). MPLUS does not print in the output the diagonal matrix $\mathbf{U} = (1-2\tilde{\sigma}^2)\mathbf{I}$. This matrix consists of the variances of the pair specific errors which account for the intransitivities observed in the data. These variances are assumed to be invariant across all paired comparisons. In these data, the pair specific variance is $1-2\tilde{\sigma}^2 = 0.08$.

Thus, according to the model the most preferred model in the population is the Opel Corsa, followed by the Volkswagen Polo, then by the Renault Clio, and the least preferred model is the Seat Ibiza. Also, the between-subject variability is invariant for all car models, with population variance 0.46. Furthermore, between-subject variabilities are independent across all car models. In other words, individual variability in the preferences for a model is independent of the individual variability in the preferences for any other model. Finally, the between-subject variability attributed to a car model presented next to another (pair specific between-subject variability) is invariant across paired comparisons, with population variance 0.08. This variability is much smaller than the preferences' variability. This reflects the fact that most respondents are transitive.

In closing this section, we notice that the size of the standard errors for the mean preferences relative to the parameter estimates suggest we could impose the following equality restrictions on the mean preferences $\mu_0 = \mu_V = 0$, and $\mu_R = \mu_S$. A model with these additional restrictions also fits the data very well. The mean corrected Satorra-Bentler statistic equals 18.80 on 19 df, p = 0.47. Hence, we can not reject the hypothesis that the mean preferences for the Opel Corsa and Volkswagen Polo are equal, nor that the mean preferences for the Renault Clio and Seat Ibiza are equal. These mean similarities may be due to a country-of-origin effect. Respondents may have a similar mean preference for the Opel Corsa and the Wolkswagen Polo because both car brands are German. Also, respondents may have a similar preference for the Renault Clio and Seat Ibiza because both car brands are from Southern Europe (France and Spain, respectively). Obviously, additional data would be needed to verify this country-of-origin hypothesis.

3.3.5 Modeling career preferences among Spanish Psychology undergraduates using rankings

A pilot study was performed to investigate career preferences among sophomore Psychology undergraduates from a Spanish university. 55 students were asked to rank these broad Psychology career areas {A = Academic, C = Clinical, E = Educational, and I = Industrial} according to their preferences. In Table 2 we provide all possible ranking patterns, their corresponding paired comparisons patterns and the frequency of each pattern observed in the sample.

Table 2. Observed frequencies of ranking patterns and their corresponding paired comparisons patterns in the career choices data

A	C	E	I	{A,C}	{A,E}	{A,I}	{C,E}	{C,I}	{E,I}	freq.
1	2	3	4	1	1	1	1	1	1	0
1	2	4	3	1	1	1	1	1	0	1
1	3	2	4	1	1	1	0	1	1	0
1	3	4	2	1	1	1	1	0	0	0
1	4	2	3	1	1	1	0	0	1	0
1	4	3	2	1	1	1	0	0	0	0
2	1	3	4	0	1	1	1	1	1	0
2	1	4	3	0	1	1	1	1	0	1
2	3	1	4	1	0	1	0	1	1	0
2	3	4	1	1	1	0	1	0	0	0
2	4	1	3	1	0	1	0	0	1	1
2	4	3	1	1	1	0	0	0	0	0
3	1	2	4	0	0	1	1	1	1	6
3	1	4	2	0	1	0	1	1	0	3
3	2	1	4	0	0	1	0	1	1	6
3	2	4	1	0	1	0	1	0	0	1
3	4	1	2	1	0	0	0	0	1	1
3	4	2	1	1	0	0	0	0	0	1
4	1	2	3	0	0	0	1	1	1	12
4	1	3	2	0	0	0	1	1	0	11
4	2	1	3	0	0	0	0	1	1	0
4	2	3	1	0	0	0	1	0	0	6
4	3	1	2	0	0	0	0	0	1	4
4	3	2	1	0	0	0	0	0	0	1

Notes: N = 55; A = Academic, C = Clinical, E = Educational, I = Industrial; in the rankings, the most preferred stimuli is assigned a 1; the binary variables take a value of 1 if the first stimuli within a pair ranked over the second stimuli and 0 otherwise.

Because we are now modeling rankings we fitted Thurstone's original Case V model (27) to the paired comparisons obtained from this ranking. Again, because the sample size was rather small we used DWLS estimation. The mean adjusted Satorra-Bentler statistic printed in the MPLUS output is 22.11 on 18 degrees of freedom. However, using (22) we obtain the correct number of degrees of freedom: 18 - 4 = 14, p = 0.08. So the model fits the data reasonably well. The parameter estimates with standard errors in parentheses are μ_A = -0.78 (0.16), μ_C = 0.71 (0.17), μ_E = 0.17 (0.17), where μ_i = 0 and σ^2 = ½ for identification purposes.

Thus, according to the model the most preferred career area is clinical, followed by educational, followed by industry, and the least preferred career area is academic. Also, the between-subject variability is invariant for all career areas, and the between-subject variability is independent across all careers. In other words, individual variability in the preferences for a career area is independent of the individual variability in the preferences for another career area.

3.4 Thurstonian Case V modeling of paired comparisons and ranking data when background information on the respondents is available

Often times background information on the respondents such as age, gender, personality characteristics, etc. is available. When no background information was included in the model, we assumed an underlying multivariate normal distribution with some threshold and covariance structure. We can incorporate respondents' background variables into the model by treating them as exogenous variables and by assuming an underlying multivariate normal distribution conditional on the background variables (Muthén, 1982, 1984). That is, we need not assume that the background variables are also multivariate normal. In this way, we can model for example the effects of binary variables such as gender on the observed preferences. Furthermore, within this conditional approach the relationships among the exogenous variables is not modeled. Only the effects of the exogenous variables (respondents' background variables) on the endogenous variables (paired comparisons or rankings) and the relationships among the endogenous variables are modeled.

In previous sections, we saw how to model the relationships among the endogenous variables using Thurstone's Case V model (for ranking data) and the Thurstone-Takane Case V model (for paired comparisons data). The restrictions imposed by the Thurstone-Takane Case V model on the thresholds and tetrachoric correlations among the variables $\mathbf{z}^*$ are given by (29), whereas the restrictions imposed by Thurstone's Case V model are given by (27). When p exogenous variables $\mathbf{x}$ are present, we also need to model the slopes of the regression of $\mathbf{x}$ on the $\tilde{n}$-dimensional vector $\mathbf{z}^*$. The restrictions imposed by the Thurstone-Takane and Thurstone's Case V models on the $\tilde{n} \times p$ matrix of slopes $\mathbf{\Pi}$ are

$$\mathbf{\Pi} = \mathbf{A\Gamma} . \tag{33}$$

Here, Γ denotes the $n \times p$ matrix of slopes of the regression of $\mathbf{x}$ on $\mathbf{t}$. Not all parameters in $\mathbf{\Gamma}$ can be identified. Because of the comparative nature of the data, to identify the model we need to fix the elements of one row of $\mathbf{\Gamma}$. We shall fix at zero the elements of the last row of $\mathbf{\Gamma}$. This implies that we are assuming that the exogenous variables have no effect on the last stimulus being compared. The remaining regression slopes must therefore be interpreted relative to this assumption. In other words, because of the comparative nature of the data, we are unable to determine if an exogenous variable has or has not an effect on the preferences for a stimulus. We can only determine if the effect of an exogenous variable on a stimulus is larger or smaller than the effect for a reference stimulus.

There is a close correspondence between these Thurstonian models with exogenous variables and the structural multivariate probit model with latent variables discussed by Muthén (1979). Also, there is a close correspondence between these Thurstonian models with exogenous variables and a MIMIC model (Jöreskog & Goldberger, 1975) with binary indicators. Again, the key to these correspondences is to view Thurstonian models as factor models with non-zero factor means where the matrix of factor loadings consists of fixed constants given by $\mathbf{A}$.

To illustrate the present discussion, we shall now return to the compact cars example and model the responses to the same car models {O = Opel Corsa, R = Renault Clio, S = Seat Ibiza, V = Volkswagen Polo} along with background information on the respondents. The full compact cars data is provided in Maydeu-Olivares (1998). Only three background variables are included in the data set: gender (coded as male = 0 and female = 1), age (ranging from 18 to 21) and monthly family income. Family income is a categorical variable with categories 'less than 100,000 Spanish pesetas', 'between 100,000 and 200,000 pesetas', 'between 200,000 and 300,000 pesetas' and 'more than 300,000 pesetas'. We coded this variable using the values {0.8, 1.5, 2.5, 3.2}. Complete observations were available for 196 respondents, of which 92 were men, and 104 women.

To model these paired comparisons data, we fitted the Thurstone-Takane Case V model (29) with (33). When exogenous variables are present MPLUS estimates first the thresholds, the regression slopes $\mathbf{\Pi}$ and the tetrachoric correlations, and collects them in a vector $\boldsymbol{\kappa}$. Then the model parameters are estimated by minimizing (32). In this final estimation stage one may use WLS or DWLS. For our example, we used DWLS. The mean corrected Satorra-Bentler statistic for this model is 49.40 on 26 df, $p < 0.01$, whereas the mean and variance corrected Satorra-Bentler statistic is 24.70 on 13 df, $p = 0.03$. Hence the model barely fits these data.

The parameter estimates obtained and their standard errors are provided in Table 3.

Table 3. Parameter estimates for a Thurstone-Takane Case V model for the compact cars data incorporating the respondents' age, gender and family income

Initial model

		regression slopes		
Car model	μ	***gender***	***age***	***family income***
Opel Corsa	-1.49 (1.54)	0.18 (0.16)	0.08 (0.07)	0.02 (0.09)
Renault Clio	-4.16 (1.73)	0.39 (0.17)	0.17 (0.09)	0.15 (0.10)
Seat Ibiza	-4.12 (1.60)	0.48 (0.16)	0.18 (0.08)	0.07 (0.10)
Volkswagen Polo	0*	0*	0*	0*

Final model

		regression slopes		
Car model	μ	***gender***	***age***	***family income***
Opel Corsa	0^*	0^*	0^*	0^*
Renault Clio	-3.40^a (1.17)	0.35^b (0.11)	0.14^c (0.06)	0^*
Seat Ibiza	-3.40^a (1.17)	0.35^b (0.11)	0.14^c (0.06)	0^*
Volkswagen Polo	0^*	0^*	0^*	0^*

Notes: $N = 196$; standard errors in parentheses; $^{a, b, c}$ denote parameters constrained to be equal to other parameters in the model; * parameter fixed for identification purposes; $\tilde{\sigma}^2 = 0.47$ (0.02)

As shown in this Table, the mean preference for the last car model, Volkswagen Polo, as well as the regression slopes of the exogenous variables on this car model were fixed at zero to identify the model. Therefore, the remaining regression slopes must be interpreted as follows: If family income has no effect on the preferences for the Volkswagen Polo, then family income has no significant effects on the preferences for any of the remaining car models. On the other hand, if age has no effect on the preferences for the Volkswagen Polo, there are significant age effects on the preferences for the Renault Clio and Seat Ibiza. Older respondents are more likely to prefer these two car models. Finally, if there are no gender effects on the preferences for the Volkswagen Polo, there are significant gender effects on the preferences for the Renault Clio and Seat Ibiza. Women are more likely to prefer these two car models.

Finally, as shown in Table 3, the most preferred car model is now Volkswagen Polo followed by Opel Corsa, then by Seat Ibiza, and finally by

Renault Clio. The standard errors for these parameters suggest that we could constrain the mean preference for the Volkswagen Polo and Opel Corsa to be equal, and that we could also constrain the mean preference for the Renault Clio and Seat Ibiza to be equal. With these constraints the ordering of the preferences for these car models is the same that the one we obtained when we model the preferences from male respondents. Actually, only four parameters seem to be needed to reproduce these data, as upon inspection of their standard errors the slopes of gender on the preferences for the Renault Clio and the Seat Ibiza can be set equal, and the slopes of age on the preferences for the Renault Clio and the Seat Ibiza can also be set equal.

A model with these constraints fits the data reasonably well. The Satorra-Bentler mean scaled statistic yields 49.40 on 26 df, $p = 0.06$, and the Satorra-Bentler mean and variance adjusted statistic yields 23.94 on 17 df, $p = 0.12$. The resulting parameter estimates and standard errors are also presented in Table 3. This final model can be interpreted as follows: Purchasing preferences for the Volkswagen Polo and the Opel Corsa appear to be the same, and so do purchasing preferences for the Renault Clio and the Seat Ibiza. Respondents prefer the Volkswagen Polo and the Opel Corsa over the Renault Clio and the Seat Ibiza. There are no differential effects of family income on purchasing preferences for any of these models. However, older respondents are more likely to prefer the Renault Clio and the Seat Ibiza, and women are more likely than men to prefer the Renault Clio and the Seat Ibiza. Furthermore, the between-subject variability in purchasing preferences is invariant across all car models, the between-subject variability in paired specific preference deviancies is also invariant across all paired comparisons, and the deviancies of individual preferences from the mean preferences are independent for all car models.

3.5 Conclusions

We have reviewed Thurstone's classical Case V model for paired comparisons and ranking data. Thurstone's Case V model is not a proper model for paired comparisons data when the same individuals respond to all paired comparisons as it assigns zero probabilities to all intransitive paired comparisons patterns. However, Thurstone's Case V model is a proper model for ranking data. To model multiple judgment paired comparisons a vector of pair specific errors needs to be added to Thurstone's model, following Takane (1987). Although the models appear cumbersome, they can be straightforwardly estimated within a structural equations approach using software capable of handling binary indicators such as MPLUS. The limited information estimation approach employed in MPLUS is a natural extension of the classical estimation procedure for the Case V model. This is discussed in the Appendix.

In this paper, we have only discussed the most widely known variant of Thurstone's model -his Case V model. This is also the most restrictive variant of Thurstone's model. If this model is found to provide a poor fit to their data applied researchers may want to consider less restrictive Thurstonian models. For instance: (a) Thurstone's Case III model in which the "discriminal processes" are assumed to be independent but with different variances, or (b) an unrestricted Thurstonian model in which only minimal identification restrictions are imposed on the mean vector and covariance matrix of the discriminal processes. Alternatively, applied researchers may be interested in Thurstonian model in which a dimensional model such as an ideal point or a factor model is assumed to underlie the discriminal processes. However, the more complex Thurstonian models impose constraints on the thresholds and tetrachoric correlations that can not be enforced using standard software for structural equation modeling at the time of this writing. Fortunately, any Thurstonian model for paired comparisons and ranking data can be fitted using the less well-known structural equations modeling package MECOSA (Arminger, Wittenberg & Schepers, 1996).

In any case, the structural equations modeling framework described here is a promising approach to estimate Thurstonian paired comparisons and ranking models. Simulation studies (Maydeu-Olivares, 2001, 2003) indicate that using this approach adequate parameters estimates, standard errors and goodness of fit tests can be obtained for Thurstonian models for 7 stimuli with as few as 100 observations. Smaller sample sizes are required to fit smaller models.

References

Arminger, G., Wittenberg, J. & Schepers, A. (1996). *MECOSA 3. User guide.* Friedrichsdorf: Additive GmbH.

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

Böckenholt, U. (1990). Multivariate Thurstonian models. *Psychometrika, 55*, 391-403

Böckenholt, U. (1993). Applications of Thurstonian models to ranking data. In M.A. Fligner and J.S. Verducci (Eds). *Probability models and statistical analyses for ranking data*. New York: Springer-Verlag.

Böckenholt, U. (2001). Hierarchical modeling of paired comparison data. *Psychological Methods*, *6*, 49-66.

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

Chan, W. & Bentler, P.M. (1998). Covariance structure analysis of ordinal ipsative data. *Psychometrika*, *63*, 360-369.

David, H.A. (1988). *The method of paired comparisons*. London: Griffin.

Jöreskog, K.G. & Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 69, 631-639.

Luce, R.D. (1959). *Individual choice behavior*. New York: Wiley.

Maydeu-Olivares, A. (1998). Structural equation modeling of binary preference data. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *58*, 5694.

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.

Maydeu-Olivares, A. (2002). Limited information estimation and testing of Thurstonian models for preference data. *Mathematical Social Sciences*, *43*, 467-483.

Maydeu-Olivares, A. (2003). On Thurstone's model for paired comparisons and ranking data. In Yanai, H., Okada, A., Shigematu, K., Kano, Y., Meulman, J.J. (Eds.). *New Developments in Psychometrics* (pp. 519-526) Tokyo: Springer.

Mosteller, F. (1951a). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, *16*, 3-9.

Mosteller, F. (1951b). Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, *16*, 207-218.

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

Muthén, B. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association*, *74*, 807-811.

Muthén, B. (1982). Some categorical response models with continuous latent variables. In K.G. Jöreskog & H. Wold (Eds.). *Systems under indirect observation*. (Vol 1) (pp. 65-79). Amsterdam: North Holland.

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. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Paper accepted for publication in *Psychometrika*.

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

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.

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

Torgerson, W.S. (1958). *Theory and methods of scaling*. New York: Wiley.

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.

Tsai, R.C. & Böckenholt, U. (2001). Maximum likelihood estimation of factor and ideal point models for paired comparison data. *Journal of Mathematical Psychology*, *45*, 795-811.

Tsai, R.C. & Yao, G. (2000). Testing Thurstonian Case V ranking models using posterior predictive checks. *British Journal of Mathematical and Statistical Psychology*, *53*, 275-292.

Yao, G. & Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology*, *52*, 79-92.

Yu, P.L.H. (2000). Bayesian analysis of order-statistics models for ranking data. *Psychometrika, 65*, 281-299.

Appendix

Relationship between classical and SEM estimation approaches to Thurstone's Case V model under multiple judgment sampling

In the notation employed in this paper, Mosteller (1951a) showed that the classical estimation solution (see Torgerson, 1958) to Thurstone's Case V model is equivalent to minimizing the unweighted least squares function

$$F = (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}(\boldsymbol{\theta}))'(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}(\boldsymbol{\theta}))\,, \tag{34}$$

where

$$\hat{\tau}_{i,j} = -\Phi^{-1}(p_{i,j}) \,, \tag{35}$$

and $p_{i,j}$ is the sample proportion corresponding to (2).

Now, the proportions $p_{i,j}$ are the sample means of the binary variables $y_{i,j}$ and $\hat{\tau}_{i,j}$ is the maximum likelihood estimate of $\tau_{i,j}$ when estimated from $y_{i,j}$. Thus, the classical estimation approach consists of a two stage procedure that only makes use of univariate information. In the first stage each threshold is estimated one variable at a time using (35) and in the second stage the model parameters are estimated by unweighted least squares from the first stage estimates.

In contrast, within a structural equations approach a three stage procedure using univariate and bivariate information is used (Muthén, 1978, 1984). In the first stage, each threshold is estimated separately one variable at a time using maximum likelihood. In the second stage, each tetrachoric correlation is estimated separately from two variables at a time inputting the first stage estimates. Finally, in the third stage the model parameters are estimated from the first and second stage estimates using either an unweighted, a diagonally weighted or a full weighted least squares approach. Therefore, the structural equations approach is a natural extension to the classical procedure in which not only univariate information from the data is employed, but also bivariate information.

Standard errors and goodness of fit tests are computed within a structural equations framework taking into account the dependencies among the univariate and bivariate proportions. In contrast, in the classical estimation approach, the univariate proportions $p_{i,j}$ are assumed to be independent. This assumption is violated in the case of multiple judgment paired comparisons and ranking data –see (15). Therefore, standard errors and goodness of fit tests for these data obtained using the classical approach are incorrect. For instance, Mosteller's (1951b) goodness of fit test is overly optimistic when applied to multiple judgment paired comparisons and ranking data.

Furthermore, many Thurstonian models are not identified from univariate information alone. Yet, any Thurstonian model can be are identified as soon as bivariate information is used. For instance, the Thurstone-Takane Case V model is not identified from univariate information alone. Therefore, it can not be estimated using the classical estimation approach for Thurstonian models.