New Developments in Psychological Choice Modeling
G. De Soete, H. Feger and K. C. Klauer (eds.)

ANALYSIS OF COVARIANCE STRUCTURES AND PROBABILISTIC BINARY CHOICE DATA

Yoshio Takane
McGill University, Canada

Pair comparison judgments are often obtained by multiple-judgment sampling, which gives rise to dependencies among observations. Analysis of covariance structures (ACOVS) provides a general methodology for taking apart between-subject and within-subject variations, thereby accounting for the dependencies among observations. In this expository paper we show how various concepts underlying ACOVS can be used in constructing probabilistic choice models that take into account systematic individual differences.

1. Introduction

Stimulus comparison presents a general paradigm in diversified fields of scientific investigations (Bradley, 1976). In bioassay strength of life of an organism is compared with dosage levels of a drug. In psychology, econometrics and political science, a subjective quality of a stimulus (e.g., subjective length of a line, grayness of a color, preference toward a political candidate, etc.) is compared against that of another. In statistics loglinear analysis of a frequency table compares the strengths with which subjects belong to certain categories. In a mental test subjects' ability is compared against difficulty of a test item.

The work reported in this paper has been supported by Grant A6394 to the author from the Natural Sciences and Engineering Research Council of Canada. Thanks are due to Jim Ramsay for his helpful comments on an earlier draft of this paper.

This paper is a revised version of an article published in *Communication & Cognition*, 1987, *20*, 45-62.

In each case, p_{ij}, the probability that stimulus i is chosen over stimulus j, indicates the degree to which stimulus i dominates stimulus j. However, there are two possible interpretations of p_{ij}, which closely parallel two sampling schemes of pair comparison data (Thurstone, 1927). In Case 1 replications (both within and across stimulus pairs) are made strictly within a single subject, and thus inconsistency in choice is attributed to momentary fluctuations in the internal state of the subject. The p_{ij} in this case represents the proportion of times stimulus i is chosen over stimulus j by the subject. In Case 2, on the other hand, the probability distribution is over a population of subjects. That is, the stochastic nature of choice is attributed to subject differences. The p_{ij} in this case represents the proportion of the subjects in the population who choose stimulus i over stimulus j.

Despite the difference in the interpretation, basically the same class of models have been used in both cases. Typically, these models assume statistical independence among observed choice probabilities. However, in Case 1 all pair comparison judgments are made by a single subject, so that there should be no sequential effects. This rules out identifiable stimuli to be used in Case 1 because of the memory effect. In Case 2, each subject is supposed to contribute one and only one observation. This usually ensures the statistical independence. On the other hand, it requires a huge number of subjects. Pair comparison experiments thus rarely use either one of these extreme sampling designs. Instead they typically employ a mixed design, in which each of a group of subjects is asked to respond to all possible pairs of stimuli. That is, replications over different stimulus pairs are obtained within subjects, and replications within stimulus pairs are obtained across subjects. This mixed mode sampling scheme is analogous to the treatment by subject design in ANOVA and is called multiple-judgment sampling in this paper. This sampling design is especially popular in preference judgments, because researchers in this area are often interested in how preferences toward various stimuli correlate with each other, how patterns of preference distribute in the population of subjects, and how an individual's pattern of preference can be represented in relation to others.

In the multiple-judgment sampling p_{ij} can be still interpreted as the proportion of the subjects who choose stimulus i over stimulus j, as in Case 2. However, due to within-subject replications across different stimulus pairs, observed choice probabilities are no longer statistically independent. Systematic individual differences give rise to the dependencies among the observations. For example, a person who tends to prefer product A to B may also tend to prefer C to D. Models of pair comparisons in this case should take into account the systematic individual differences in pair comparison judgments. However, with notable exceptions (Bock & Jones, 1968, pp. 143-161; Bloxom, 1972; Takane, 1985) nearly all previous models of pair comparisons ignored the systematic individual differences.

What is needed is a general methodology for separating the systematic individual differences components in the data from strictly random components. The method particularly relevant in this context is the analysis of covariance structures (ACOVS) originally proposed by Bock and Bargman (1966) and subsequently amplified by Jöreskog (1970). As has been demonstrated recently (Takane, 1985), the ACOVS framework can be successfully used to extend conventional Thurstonian pair comparison models to multiple-judgment sampling situations. In addition the ACOVS framework may bring on considerable richness to analysis of pair comparison data in general. The purpose of this paper is to explore and overview this possibility.

2. Thurstonian Models of Pair Comparisons

Let us begin with a brief review of Thurstonian random utility models (Thurstone, 1927, 1959). Over the past several years there were interesting developments in this approach (Takane, 1980; Heiser & de Leeuw, 1981; Carroll, 1980; De Soete & Carroll, 1983), which directly lead to the ACOVS formulations of these models.

In Thurstone's original pair comparison model each stimulus is associated with a random variable (called a discriminal process) with prescribed distributional properties. Let Y_i represent the random variable for stimulus i. It is assumed that

$$Y_i \sim N(m_i, \sigma_i^2), \quad i = 1, ..., n \tag{1}$$

where $m_i = E(Y_i)$ and $s_i^2 = V(Y_i)$. The m_i represents the mean scale value (e.g., preference value), and s_i^2 the degree of uncertainty of stimulus i. When stimuli i and j are presented for comparison, random variables corresponding to these stimuli, namely Y_i and Y_j, are generated, and the comparison is supposedly made on the realized values of the random variables at the particular time. The comparison process is supposed to take the difference between Y_i and Y_j, and either the value of $Y_i - Y_j$ or some monotonic transformation of it is directly reported, or only its sign (if $Y_i - Y_j$ is positive or negative) is reported in the form of choice (either stimulus i is chosen or stimulus j is chosen). Under the distributional assumption made above,

$$Y_i - Y_j \sim N(m_i - m_j, d_{ij}^2), \tag{2}$$

where

$$d_{ij}^2 = V(Y_i - Y_j) = s_i^2 + s_j^2 - 2s_{ij} \tag{3}$$

with $s_{ij} = \mathrm{Cov}(Y_i, Y_j)$. Thus the probability that stimulus i is chosen over stimulus j is given by

$$\begin{aligned} p_{ij} &= \Pr(Y_i - Y_j > 0) \\ &= \int_{-\infty}^{q_{ij}} \phi(z)\mathrm{d}z = \Phi(q_{ij}) \end{aligned} \tag{4}$$

where $q_{ij} = (m_i - m_j)/d_{ij}$, and ϕ and Φ are, respectively, the density function and the cumulative distribution function of the standard normal distribution. The d_{ij} indicates the degrees of uncertainty in the comparison. When the uncertainty is small, even a small difference between the m_i and m_j makes a lot of difference in the choice probability. If, on the other hand, there is a great deal on uncertainty in the comparison, the choice probability is relatively insensitive to the difference between m_i and m_j. The d_{ij} is sometimes called an uncomparability index (Halff, 1976).

Despite its generality and appeal Thurstone's general pair comparison model (4) has one major drawback. The number of parameters in the model exceeds the number of observed binary choice probabilities. Some simplifying assumption is therefore necessary. In the simplest possible case it is assumed that $d_{ij} = 1$ for all i and j (Case 5). However, this implies that all stimulus pairs are equally comparable. It also implies the context independence of the pair comparison process. That is, the choice probability is a function of only scale values of the stimuli involved, and these scale values remain invariant no matter with which stimuli the particular stimuli are compared. Krantz (1967) calls this condition "simple scalability". However, numerous studies (Debreu, 1960; Krantz, 1967; Restle, 1961; Tversky & Russo, 1969; Rumelhart & Greeno, 1970; Tversky, 1972a, b; Sjöberg, 1977, 1980) reported violations of simple scalability in a variety of empirical situations.

All stimuli are not equally comparable. The equal comparability holds only when stimuli to be compared are relatively homogeneous. When the stimuli are radically different on "irrelevant" dimensions (i.e., dimensions other than the one on which the comparison is supposedly made), they tend to be less comparable, and the choice probabilities tend to be less extreme (closer to 1/2). If, on the other hand, the stimuli are similar, they are more comparable, and consequently more extreme choice probabilities tend to result (Krantz, 1967; Tversky & Russo, 1969; Rumelhart & Greeno, 1971). Thus differential degrees of similarity among stimuli give rise to context dependencies in the stimulus comparison process, called the similarity effect.

This means that d_{ij} in Thurstone's original model has its role to play. In particular, it has been shown (Halff, 1976) that d_{ij} has distance properties, and d_{ij} satisfies the three metric axioms (minimality, symmetry and triangular inequality) required of the distance. The distance properties of d_{ij} make Thurstone's general model considerably richer in its descriptive power than those models that assume simple scalability. Specifically, Thurstone's general pair comparison model satisfies moderate stochastic transitivity (MST), but it can violate strong stochastic transitivity (SST), which is known to be equivalent to the simple scalability (Tversky & Russo, 1969).

It is interesting to point out that d_{ij}, the distance between stimuli i and j, can be interpreted as a type of dissimilarity between the stimuli. Thus, dividing, $m_i - m_j$ by d_{ij} in q_{ij} in Thurstone's general model is consistent with the empirical evidence (mentioned earlier) indicating that more dissimilar stimuli are less comparable. Sjöberg (1977) observed a high correlation between d_{ij} estimated from pair comparison judgments and a direct similarity rating between stimuli i and j separately obtained. The d_{ij} is thus not only theoretically expected to represent the stimulus dissimilarity, but there is also some empirical evidence to support the theory.

The problem is how we may recover d_{ij} in Thurstone's general model without overparametrizing it. Attempts to extend Thurstone's pair comparison model beyond Case 5 are almost as old as Thurstone's original proposal of the model (Thurstone, 1927). For example, in Case 3 it is assumed that $s_{ij} = 0$ for all i and j, thereby reducing the number of parameters considerably. Case 4 was derived as a convenient numerical approximation to Case 3. However, in these cases differential comparability (d_{ij}) between stimuli is exclusively attributed to individual uncertainties (s_i^2 and s_j^2). Thus, they are rather restrictive as models of contextual effects in stimulus comparison processes.

A couple of significant proposals were made in early 1980's in the way of partially recovering d_{ij} in Thurstone's model. Takane (1980) and Heiser and de Leeuw (1981) independently proposed the factorial model of pair comparisons (hereafter called the THL model), in which the covariance matrix between discriminal processes was assumed to have a lower rank approximation. That is,

$$\mathbf{S} = (s_{ij}) = \mathbf{XX}' \tag{5}$$

where X is an n by b ($< n$) matrix where n is the number of stimuli and b is the rank of matrix S. This amounts to assuming

$$d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j), \tag{6}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are i th and j *th* row vectors of $\mathbf{X}$, since $s_i^2 = \mathbf{x}_i{}'\mathbf{x}_j$ and $s_j^2 = \mathbf{x}_j{}'\mathbf{x}_j$. That is, d_{ij} is assumed to be the Euclidean distance between stimuli i and j represented in a b dimensional Euclidean space. The $\mathbf{X}$ then represents the matrix of stimulus coordinates.

An interesting development was due to Carroll (1980) and De Soete and Carroll (1983). The model is called the wandering vector model (WVM). In this model it is assumed that stimuli are represented as points in a b dimensional space where stimulus coordinates are given by $\mathbf{X}$ as in the THL model, that there is a random vector that varies over time, and that the projections of the stimuli onto this vector at a particular time determines the pair comparison judgment at the particular time. Under an appropriate distributional assumption on the vector we may derive the distribution of $Y_i - Y_j$, and the choice probability, p_{ij}. Let $\mathbf{u}^*$ denote the wandering vector, and let $\mathbf{u}^* \sim N(\mathbf{v},\mathbf{I})$. Then

$$Y_i - Y_j = (\mathbf{x}_i - \mathbf{x}_j)'\mathbf{u}^* \sim N[(\mathbf{x}_i - \mathbf{x}_j)'\mathbf{v},\ d_{ij}^2], \tag{7}$$

where d_{ij} is the same as in (6). It follows that

$$p_{ij} = \Pr[(\mathbf{x}_i - \mathbf{x}_j)'\mathbf{u}^* > 0]$$
$$= \int_{-\infty}^{r_{ij}} \phi(z)dz = \Phi(r_{ij}), \qquad (8)$$

where $r_{ij} = (\mathbf{x}_i - \mathbf{x}_j)'\mathbf{v}/d_{ij}$.

It has been shown (De Soete, 1983) that the WVM is a special case of the THL model in which not only d_{ij} but also m_i and m_j are constrained in a special way; i.e.,

$$m_i = \mathbf{x}_i'\mathbf{v} \text{ and } m_j = \mathbf{x}_j'\mathbf{v}. \qquad (9)$$

Scale values of stimuli are represented in a particular direction in the space. Thus, although the THL model and the WVM were initially derived on the basis of entirely different rationales, they are quite similar to each other.

Both the THL model and the WVM are designed to account for the differential comparability among the stimuli. However, these models strictly apply to either Case 1 or Case 2, where differences processes, $Y_i - Y_j$, and consequently observed choice probabilities, are assumed statistically independent across all pairs of stimuli. Both Takane (1980) and De Soete and Carroll (1983) developed parameter estimation procedures for their models. They both assume the statistical independence among the observations, while they use the data obtained by the multiple-judgment sampling. As has been discussed, the independence assumption is not tenable in the multiple judgment sampling. However, the assumption is made in virtually all previous estimation procedures for the Thurstonian pair comparisons models (e.g., Hohle, 1966; Bock & Jones, 1968; Arbuckle & Nugent, 1973; Takane, 1980; De Soete & Carroll, 1983; De Soete, Carroll, & DeSarbo, 1986). In order to account for the statistical dependencies among observations, pair comparison models had to await analysis of covariance structure formulations (Bloxom, 1972; Takane, 1985), to which we now turn.

In closing of this section it might be pointed out that analogous developments (models of simple scalability to moderate utility models) can be traced in the Bradley-Terry-Luce (Bradley & Terry, 1952; Luce, 1959) type of constant utility model approach (Restle, 1961; Tversky, 1972a, b; Strauss, 1981). However, these developments are not readily amenable to the ACOVS formulations. See Indow (1975) and Luce

(1977) for insightful reviews of this line of development.

3. ACOVS Formulations

In order to reformulate the THL model and the WVM in terms of analysis of covariance structures (ACOVS; Jöreskog, 1970), let us first generalize the variance structure of these models. It was originally assumed that $d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)$ in these models. To this we may add $g_i^2 + g_j^2 + k_{ij}^2$, where g_i^2 and g_j^2 are stimulus-specific uncertainties left unaccounted for by $(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)$, and k_{ij}^2 represents uncertainty associated with a specific stimulus pair. These quantities represent amounts of specification error at two different levels.

We now generalize this to covariance structures. Let $\mathbf{t}$ be a vector of $t_{ij} = Y_i - Y_j + e_{ij}$ arranged in a specific order, where e_{ij} is the error random variable associated with stimulus pair, ij. In a complete sampling design each subject makes judgments for all possible pairs of stimuli. In such a case $\mathbf{t}$ is a of dimensionality $M = n(n-1)/2$, where n is the number of stimuli. Let $\mathbf{A}$ be an M by n design matrix for pair comparisons, whose rows are arranged in the same order as the elements of $\mathbf{t}$. Each row of $\mathbf{A}$ corresponds with a specific comparison. If that comparison involves stimuli i and j and the direction of the comparison requires $Y_i - Y_j$ (rather than $Y_j - Y_i$), the row has 1 in the ith column, -1 in the jth column and zeroes elsewhere.

Let $\mathbf{y}$ be an n-component vector of Y_i, and let $\mathbf{e}$ be an M-component vector of e_{ij}. We assume

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{K}^2) \tag{10}$$

where $\mathbf{K}^2$ is assumed to be diagonal with its diagonal elements denoted by k_{ij}^2. It may be further assumed $k_{ij}^2 = k^2$ for all ij. Then $\mathbf{t}$ may be expressed, using matrix notation, as

$$\mathbf{t} = \mathbf{A}\mathbf{y} + \mathbf{e}. \tag{11}$$

The $\mathbf{A}\mathbf{y}$ takes differences between Y_i and Y_j in prescribed directions for all possible pairs of stimuli.

We make a further structural assumption on $\mathbf{y}$; namely,

$$\mathbf{y} = \mathbf{X}\mathbf{u}^* + \mathbf{w}^*, \tag{12}$$

where

$$\mathbf{w}^* = \mathbf{w} + \mathbf{m} \quad \text{with} \quad \mathbf{w} \sim N(\mathbf{0},\mathbf{G}^2), \tag{13}$$

and

$$\mathbf{u}^* = \mathbf{u} + \mathbf{v} \quad \text{with} \quad \mathbf{u} \sim N(\mathbf{0},\mathbf{I}). \tag{14}$$

Here $\mathbf{m}$ is the vector of m_i ($i = 1,..., n$) and $\mathbf{w}$ is the random vector of stimulus specificities. The $\mathbf{G}^2$ is usually assumed to be diagonal with its ith diagonal element, g_i^2, and indicates the degrees of stimulus specificities or uncertainties. The $\mathbf{u}^*$ is the wandering vector introduced earlier.

It follows that

$$\begin{aligned} \mathbf{t} &= \mathbf{A}(\mathbf{X}\mathbf{u}^* + \mathbf{w}^*) + \mathbf{e} \\ &\sim N[\mathbf{A}(\mathbf{X}\mathbf{v} + \mathbf{m}),\ \mathbf{A}(\mathbf{X}\mathbf{X}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2]. \end{aligned} \tag{15}$$

When it is assumed that $\mathbf{v} = \mathbf{0}$, then $E(\mathbf{t}) = \mathbf{Am}$, and since $\mathbf{Am}$ is the vector of $m_i - m_j$, this case corresponds with the THL model. If, on the other hand, it is assumed that $\mathbf{m} = \mathbf{0}$ we obtain $E(\mathbf{t}) = \mathbf{AXv}$. This represents the mean structure, $(\mathbf{x}_i - \mathbf{x}_j)'\mathbf{v}$, required of the WVM. The covariance structure, $\mathbf{A}(\mathbf{XX}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2$, remains the same for the both models. Note that diagonals of this covariance matrix are of the form, $(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j) + g_i^2 + g_j^2 + k_{ij}^2$, which is indeed the variance structure required of both the THL model and the WVM. Note also that off-diagonal elements of $\mathbf{A}(\mathbf{XX}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2$ are no longer zero, implying non-independence among the elements of $\mathbf{t}$. It is interesting to note that the WVM is a random effect alternative to Bechtel, Tucker, and Chang's (1971) scalar product model. In this model subjects are treated as fixed effects; i.e., for subject k, $\mathbf{t}_k = \mathbf{AXv}_k$ and $\mathbf{v}_k$ is explicitly estimated for each k.

Analogous ACOVS formulations of classical Case 5 and Case 3 are also possible. Although these cases are not likely to provide satisfactory descriptions of pair comparison data, they may serve as good benchmark models. In Case 5 d_{ij} is assumed to be constant across all combinations of i and j. The simplest way this could occur is when s_i^2 and s_j^2 are constant, and s_{ij} is zero. In the ACOVS formulation this can be achieved by setting $\mathbf{X} = \mathbf{0}$, and $\mathbf{G}^2 = s^2\mathbf{I}$. Note that $s_{ij} = 0$ is not absolutely necessary to achieve d_{ij} = constant. It is sufficient to have s_{ij} = constant (Guttman, 1954). This case corresponds with $\mathbf{X} = \mathbf{A}\mathbf{1}_n$ where $\mathbf{1}_n$ is an n-component vector of ones. However, this reduces to the previous case, since $\mathbf{A}\mathbf{1}_n = \mathbf{0}_M$. Bock and Jones (1968), in their primitive attempt to incorporate systematic individual differences in Thurstone's pair comparison model, present a model which is essentially equivalent to the ACOVS formulation of Case 5 in which $\mathbf{K}^2 = \mathbf{0}$ is also assumed. In Case 3 it is assumed that $s_{ij} = 0$ for all distinct pairs of i and j. This case can be obtained by $\mathbf{X} = \mathbf{0}$ or $\mathbf{X} = \mathbf{A}\mathbf{1}_n$, and $\mathbf{G}^2$ being diagonal (not necessarily constant).

Model (15) may be fitted to the data by the maximum likelihood or the generalized least squares method (Browne, 1974, 1984), when $\mathbf{t}$ is directly observed. In either case some existing programs, such as LISREL (Jöreskog & Sorbom, 1981), EQS (Bentler, 1985) and COSAN (McDonald, 1980), may be used for actual computation. When only choices are observed, $\mathbf{t}$ has to be reduced to choice patterns. Correspondingly the distribution of $\mathbf{t}$ must be converted into the probability distribution of the choice pattern. Let $\mathbf{h}$ denote an observed pattern, and let f be the density function of $\mathbf{t}$. Then

$$\Pr(\mathbf{h}) = \int_R f(\mathbf{t})\mathrm{d}\mathbf{t} \tag{16}$$

where R is the multidimensional rectangular region formed by the direct product of intervals R_{ij}, where $R_{ij} = (0, \infty)$ if stimulus i is chosen over stimulus j ($t_{ij} > 0$) and $R_{ij} = (-\infty, 0)$ is stimulus j is chosen over stimulus i ($t_{ij} < 0$).

Equation (16) is generally extremely difficult to evaluate due to nonzero covariances among the elements of $\mathbf{t}$. However, the first and the second order marginal probabilities are relatively easily evaluated:

$$p_{ij} = \int_0^\infty f_{ij}(t_{ij})\mathrm{d}t_{ij}, \tag{17}$$

where f_{ij} is the univariate marginal density of $t_{ij} \sim N[(m_i - m_j), (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j) + g_i^2 + g_j^2 + k_{ij}^2)]$. (The $m_i - m_j$ must be replaced by $(\mathbf{x}_i - \mathbf{x}_j)'\mathbf{v}$ in the WVM.) Similarly,

$$p_{ij,qr} = \Pr(i \text{ is chosen over } j \text{ and } q \text{ is chosen over } r)$$
$$= \int_0^\infty \int_0^\infty f_{ij,qr}(t_{ij}, t_{qr}) \mathrm{d}t_{ij} \, \mathrm{d}t_{qr}, \tag{18}$$

where $f_{ij,qr}$ is the bivariate marginal density of t_{ij} and t_{qr}. Muthen (1984) developed LISCOMP, a computer program for the generalized least squares estimation of the ACOVS model for categorical data using the first and the second order marginal probabilities. It has been shown (Christoffersson, 1975; Muthen, 1975) that a loss of information incurred by ignoring higher order marginal probabilities in the estimation is relatively minor. Alternatively, LISREL may be used with tetrachoric correlations, but it only allows the simple least squares estimation.

The ACOVS formulation of the WVM can be readily extended to the wandering ideal point (WIP) model recently proposed by De Soete et al. (1986). In the WIP model a subject is represented as a point which varies over time. The relative distances between stimulus points and the subject point at a particular time are supposed to determine preference relations observed at the particular time. The distribution of the subject point is assumed due to time-sampling of observations within a single subject. However, with the ACOVS formulation the model can be extended to the distribution of the ideal point over a population of subjects.

Let $\mathbf{u}^*$ be a random vector of coordinates of the subject point, and let

$$\mathbf{u}^* \sim N(\mathbf{v}, \mathbf{D}^2),$$

where $\mathbf{D}^2$ is a diagonal matrix. (The $\mathbf{D}^2$ can be always made diagonal by rotating the space appropriately.) Let $\mathbf{d}(\mathbf{u}^*)$ be a vector of one half times squared Euclidean distances between stimulus points and the ideal point, i.e.,

$$\mathbf{d}(\mathbf{u}^*) = 1/2 \begin{bmatrix} d_1^2(\mathbf{u}^*) \\ \cdot \\ \cdot \\ \cdot \\ d_n^2(\mathbf{u}^*) \end{bmatrix} \tag{19}$$

where $d_i^2(\mathbf{u}^*) = (\mathbf{x}_i - \mathbf{u}^*)'(\mathbf{x}_i - \mathbf{u}^*)$. In the WIP model the distance is assumed inversely related to preference. Thus, we may set

$$\mathbf{y} = -\mathbf{d}(\mathbf{u}^*) + \mathbf{w} \tag{20}$$

in (11), where $\mathbf{w}$ is defined in (13). Then

$$\begin{aligned} \mathbf{t} &= \mathbf{A}(-\mathbf{d}(\mathbf{u}^*) + \mathbf{w}) + \mathbf{e} \\ &= \mathbf{A}(\mathbf{X}\mathbf{u}^* - \tfrac{1}{2}\mathbf{x}^{(2)} + \mathbf{w}) + \mathbf{e} \\ &\sim N[\mathbf{A}(\mathbf{X}\mathbf{v} - \tfrac{1}{2}\mathbf{x}^{(2)},\ \mathbf{A}(\mathbf{X}\mathbf{D}^2\mathbf{X}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2], \end{aligned} \tag{21}$$

where

$$\mathbf{x}^{(2)} = \mathrm{diag}(\mathbf{X}\mathbf{X}')\mathbf{1}_n = \begin{bmatrix} \mathbf{x}_1'\mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n'\mathbf{x}_n \end{bmatrix}.$$

Note that this model differs from the WVM in that it has the additional $-\tfrac{1}{2}\mathbf{A}\mathbf{x}^{(2)}$ term in the mean structure $\mathbf{X}\mathbf{D}^2\mathbf{X}'$ (rather than $\mathbf{X}\mathbf{X}'$) in the covariance structure. Reparametrization by $\mathbf{X}^* = \mathbf{X}\mathbf{D}$ will make the covariance structure identical in form to that of the THL model and the WVM. However, the mean will then be $\mathbf{A}(\mathbf{X}^*\mathbf{v}^* - \tfrac{1}{2}\mathrm{diag}(\mathbf{X}^{*\prime}\mathbf{D}^{-2}\mathbf{X}^*)\mathbf{1}_n)$, so that we cannot get rid of $\mathbf{D}^2$ entirely. Vector $-\mathbf{A}\mathbf{x}^{(2)}$ has $(\mathbf{x}_j'\mathbf{x}_j - \mathbf{x}_i'\mathbf{x}_i)$ as its elements. Due to the nonlinear nature of this term, a special computer program is necessary to fit the ACOVS WIP model. An extension to choice data may be done in a manner similar to that in the WVM.

4. Possible Generalizations

A general method for analysis of covariance and mean structures (ACOVS with structured means) was given by Jöreskog (1970). The method includes, among other things, conventional factor analysis, variance-component models, path analysis, linear structural equations, etc. Our approach is a special case of this general approach. Sorbom (1981) has shown how the ACOVS with structured means could be treated in a unified manner by analysis of moment structures (AMOMS) (see also Bentler, 1983). In our case the mean and covariance structures in (15) can be expressed as

$$\mathbf{M} = \mathbf{A}(\mathbf{X}(\mathbf{v}\mathbf{v}' + \mathbf{I})\mathbf{X}' + \mathbf{m}\mathbf{m}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2 \tag{22}$$

in terms of AMOMS, where it is further assumed that $\mathbf{v} = \mathbf{0}$ or $\mathbf{M} = \mathbf{0}$.

Perhaps Bloxom (1972) was the first to note the importance of the ACOVS methodology in modeling pair comparison data. He developed his simplex model of pair comparisons (similar to Case 5) based on the ACOVS framework. Takane (1985), in an attempt to incorporate systematic individual differences into the THL model and the WVM, arrived at the ACOVS formulations of these models, which are similar in form to Bloxom's simplex model.

Working in the general ACOVS framework opens up an number of possibilities. First of all, a variety of interesting hypotheses (assumptions) can be tested explicitly. For example, $\mathbf{G}^2 = a^2\mathbf{I}$ and/or $\mathbf{K}^2 = b^2\mathbf{I}$ may be assumed and tested, or $\mathbf{G}^2 = \mathbf{0}$ and/or $\mathbf{K}^2 = \mathbf{0}$ may be assumed in (15) and their empirical validity tested. Bechtel et al.'s (1971) model corresponds with $\mathbf{m} = \mathbf{0}$, $\mathbf{G}^2 = \mathbf{0}$ and $\mathbf{K}^2 = \mathbf{0}$. In the THL model we may relax $\mathbf{XX}' + \mathbf{G}^2$ into a general positive definite matrix, $\mathbf{S}$. We then have

$$E(\mathbf{t}) = \mathbf{Am}$$

and

$$V(\mathbf{t}) = \mathbf{ASA}' + \mathbf{K}^2. \tag{23}$$

The goodness of fit comparison between this model and the original THL model tests the adequacy of the factorial decomposition of $\mathbf{S}$ into $\mathbf{XX}' + \mathbf{G}^2$.

Two particularly interesting possibilities emerge, when stimulus information and/or subject information is available. Stimuli can be characterized by a set of externally supplied attribute values (Bock & Jones, 1968), by a set of features (Rumelhart & Greeno, 1971; Tversky & Sattath, 1979), or by a set of combinations of levels of manipulated factors (Sjöberg, 1975). Similarly, subjects performing the comparisons may be characterized by their background variables, such as sex, age, socioeconomic status, levels of education, etc. In the ACOVS framework these external variables can be incorporated in a relatively straightforward manner.

Let $\mathbf{B}$ be an n by p $(< n)$ matrix of stimulus information. There are at least of couple of ways to incorporate this information. For example, we have

$$\mathbf{t} = \mathbf{A}(\mathbf{B}\mathbf{s}^* + \mathbf{X}\mathbf{u}^* + \mathbf{w}) + \mathbf{e}$$
$$\sim N[\mathbf{A}(\mathbf{B}\mathbf{m}^* + \mathbf{X}\mathbf{v}),\ \mathbf{A}(\mathbf{B}\mathbf{D}^2\mathbf{B}' + \mathbf{X}\mathbf{X}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2] \qquad (24)$$

where $\mathbf{s}^* \sim N(\mathbf{m}^*, \mathbf{D}^2)$. This model attempts to explain part of stimulus variability by **B** and the rest by **X**. This is analogous to Yanai's (1970) approach to factor analysis with external criteria, in which whatever effects that can be explained by the external criteria are first partialed out, and factor analysis is applied to a residual covariance matrix. This is to see if there is anything interesting left unaccounted for by the external criteria. More simplified or complicated versions of this model may be obtained, as desired, by specializing $\mathbf{s}^*$ in (24); e.g., $\mathbf{s}^* = \mathbf{m}^*$, $\mathbf{s}^* = \mathbf{P}\mathbf{q}^* + \mathbf{r}$, etc. In either case it may be further assumed that $\mathbf{v} = \mathbf{0}$ and/or $\mathbf{X} = \mathbf{0}$.

An alternative way to incorporate **B** is to constrain **X** by **BQ**, where **Q** is analogous to regression coefficients. This amounts to assuming that all that has been explained by **X** can be explained by **B**. We then have

$$\mathbf{t} = \mathbf{A}(\mathbf{B}\mathbf{Q}\mathbf{u}^* + \mathbf{w}) + \mathbf{e}$$
$$\sim N[\mathbf{A}\mathbf{B}\mathbf{Q}\mathbf{v},\ \mathbf{A}(\mathbf{B}\mathbf{Q}\mathbf{Q}'\mathbf{B}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2]. \qquad (25)$$

A slight generalization of this model would replace $\mathbf{Q}\mathbf{u}^*$ by $\mathbf{Q}\mathbf{u}^* + \mathbf{s}$ where $\mathbf{s} \sim N(\mathbf{0}, \mathbf{D}^2)$. We then have

$$V(\mathbf{t}) = \mathbf{A}(\mathbf{B}(\mathbf{Q}\mathbf{Q}' + \mathbf{D}^2)\mathbf{B}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2.$$

Subject information may also be incorporated in several ways. When the information is provided in nominal variables (e.g., male or female), one possibility is to partition the data into groups and to analyze them separately (Jöreskog, 1971; Muthen & Christoffersson, 1981). This allows completely different covariance structures as well as mean structures across the groups. Of course, it is entirely permissible to constrain some elements in the covariance and mean structures to be equal across the groups. In fact, the gist of the general ACOVS method is that we may explicitly test the empirical validity of such constraints.

Alternatively, subject information may be incorporated in a manner similar to regression analysis. Let $\mathbf{z}_k$ be the q-component vector of the $k$tk subject's background variables, and let $\mathbf{m}_k$ and $\mathbf{v}_k$ represent **m** and **v** in the THL model and the WVM, respectively, for subject k. We have two options. We may impose a regression structure on either $\mathbf{m}_k$ or $\mathbf{v}_k$. In the first case, we have $\mathbf{m}_k = \mathbf{Pz}_k$ and assume $\mathbf{v}_k = \mathbf{0}$, so that $E(\mathbf{t}_k) = \mathbf{APz}_k$ and $V(\mathbf{t}_k) = \mathbf{A}(\mathbf{XX}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2$ or $\mathbf{A}(\mathbf{PD}^2\mathbf{P}' + \mathbf{XX}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2$. (In either case $\mathbf{XX}' + \mathbf{G}^2$ may be replaced by a more general positive definite matrix, $\mathbf{S}$.) In the second case we assume $\mathbf{v}_k = \mathbf{P}^*\mathbf{z}_k$ while $\mathbf{m}_k = \mathbf{0}$, so that $E(\mathbf{t}_k) = \mathbf{AXP}^*\mathbf{z}_k$ and $V(\mathbf{t}_k) = \mathbf{A}(\mathbf{XX}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2$ or $\mathbf{A}(\mathbf{X}(\mathbf{P}^*\mathbf{D}^{*2}\mathbf{P}^{*\prime} + \mathbf{I})\mathbf{X}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2)$. (Again $\mathbf{XX}' + \mathbf{G}^2$ may be replaced by $\mathbf{S}$.)

Both stimulus and subject information can be simultaneously incorporated. Resulting models are combinations of those for the stimulus information and those for the subject information.

All the generalizations discussed in this section carry over to the WIP model in a relatively straightforward manner. Assuming that we have both stimulus and subject information, $\mathbf{X} = \mathbf{BQ}$ and $\mathbf{v}_k = \mathbf{P}^*\mathbf{z}_k$, we obtain, in the simplest case,

$$E(\mathbf{t}_k) = \mathbf{A}(\mathbf{BQP}^*\mathbf{z}_k - \tfrac{1}{2}\,\mathrm{diag}(\mathbf{BQQ}'\mathbf{B}')\mathbf{1}_n),$$

with $V(\mathbf{t}_k) = \mathbf{A}(\mathbf{BQQ}'\mathbf{B}' + \mathbf{G}^2)\mathbf{A}' + \mathbf{K}^2$.

5. Concluding Remarks

In this paper we have shown that the ACOVS methodology is useful in probabilistic pair comparison modeling. No empirical examples are given, and the paper largely remained expository. An obvious follow-up is to exemplify the methodological ideas described in this paper through the analyses of actual data sets. Although some of the ACOVS models for pair comparisons presented in this paper can be fitted by existing programs (e.g., LISREL, LISCOMP), there are others that cannot. For example, no ready-made programs exist for parameter estimation for the ACOVS wandering ideal point models.

The normality assumption on $\mathbf{u}$ and $\mathbf{w}$, and consequently on $\mathbf{t}$ in (15), may not be adequate. In that case we may either transform the data or use a fitting criterion that does not assume normality. Asymptotically distribution free methods (Browne, 1984) may be useful in this context.

It may appear that the proposed ACOVS models of pair comparisons have too many parameters to be estimated, particularly when the observed data are binary choices. This is indeed true for the general ACOVS model. However, it is not true in our practical applications of the ACOVS model, since matrix $\mathbf{A}$ is always a fixed matrix in the pair comparison models. The number of parameters can be further reduced, if desired, by assuming that $\mathbf{G}^2$ and/or $\mathbf{K}^2$ are constant diagonal matrices.

There are other possible generalizations that have not been explicitly discussed in this paper. For example, an extension to multiple choice situations seem to be rather straightforward. Also, treating subject's background variables as random effects (rather than fixed effects) is already feasible in LISREL (Jöreskog & Sorbom, 1981). This case corresponds with the error-in-variable regression analysis in the ACOVS framework. Our prospect of further developing the ACOVS methodology in connection with probabilistic choice models is thus bright, despite the fact that there are numerous tasks yet to be accomplished.

Appendix

How the ACOVS THL model and the WVM (15) may be fitted by LISREL is not so trivial. In this appendix we explain how this is done. We also explain how (24) and (25) can be fitted by LISREL. McArdle and McDonald (1984) provide a general framework for establishing the necessary correspondence. We appreciate Michael Browne's help (personal communication) in clarifying the matter.

The LISREL model consists of three submodels:

1. Structural Equation Model: $\tilde{\eta} = \mathbf{B}\tilde{\eta} + \Gamma\tilde{\xi} + \tilde{\zeta}$
2. Measurement Model for $\tilde{\mathbf{y}}$: $\tilde{\mathbf{y}} = \Lambda_y\tilde{\eta} + \tilde{\varepsilon}$
3. Measurement Model for $\tilde{\mathbf{x}}$: $\tilde{\mathbf{x}} = \Lambda_x\tilde{\xi} + \tilde{\delta}$,

where the symbols with a tilde on top denote random vectors. Aside from its distributional assumption (i.e., multivariate normality) the model is completely specified by the following eight matrices: Λ_y, Λ_x, $\mathbf{B}$, Γ $\Phi = E(\tilde{\xi}\tilde{\xi}')$, $\Psi = V(\tilde{\zeta})$, $\Theta_\varepsilon = V(\tilde{\varepsilon})$ and $\Theta_\delta = V(\tilde{\delta})$. (We stick with the notational convention used by Jöreskog and Sorbom (1981) as much as possible.) Throughout this appendix it is assumed that $\Lambda_x = \mathbf{I}$, $\Phi = \mathbf{I}$, $\Theta_\varepsilon = \mathbf{K}^2$ (diagonal matrix) and $\Theta_\delta = \mathbf{0}$ (zero matrix). The moment structure of $\tilde{\mathbf{y}}$ is then expressed as

$$\mathbf{M} = \Lambda_\mathbf{y}(\mathbf{I} - \mathbf{B})^{-1}(\Gamma\Gamma' + \Psi)(\mathbf{I} - \mathbf{B})^{-1}\Lambda_\mathbf{y}' + \mathbf{K}^2. \tag{A-1}$$

The following results hold:

Result 1. The moment structure of $\tilde{\mathbf{y}}$ ($\mathbf{t}$ in our notation) for the ACOVS THL model or the WVM is obtained by setting

$$\Lambda_\mathbf{y} = [\mathbf{A} \quad \mathbf{0}]$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \mathbf{m} & \mathbf{0} \\ \mathbf{0} & \mathbf{v} \end{bmatrix}$$

and

$$\Psi = \begin{bmatrix} \mathbf{G}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

(Proof) $\Lambda_\mathbf{y}(\mathbf{I} - \mathbf{B})^{-1} = [\mathbf{A} \quad \mathbf{AX}]$. Thus, (A-1) becomes

$$\mathbf{M} = [\mathbf{A} \quad \mathbf{AX}]\left(\begin{bmatrix} \mathbf{mm}' & \mathbf{0} \\ \mathbf{0} & \mathbf{vv}' \end{bmatrix} + \begin{bmatrix} \mathbf{G}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right)\begin{bmatrix} \mathbf{A}' \\ \mathbf{X}'\mathbf{A}' \end{bmatrix} + \mathbf{K}^2$$

$$= \mathbf{A}(\mathbf{mm}' + \mathbf{G}^2 + \mathbf{X}(\mathbf{vv}' + \mathbf{I})\mathbf{X}')\mathbf{A}' + \mathbf{K}^2,$$

which is identical to (22).

Result 2. The moment structure of $\tilde{\mathbf{y}}$ ($\mathbf{t}$ in our notation) corresponding to (24) is obtained by

$$\Lambda_\mathbf{y} = [\mathbf{A} \quad \mathbf{AB}^* \quad \mathbf{0}]$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{X} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{m}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{v} \end{bmatrix}$$

and

$$\Psi = \begin{bmatrix} \mathbf{G}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix},$$

where in order to avoid confusion our **B** is denoted by $\mathbf{B}^*$. (Note that in the above both **A** and $\mathbf{B}^*$ are assumed known *a priori*, so that $\mathbf{AB}^*$ can be evaluated *a priori*.)
(Proof) $\Lambda_y(\mathbf{I} - \mathbf{B})^{-1} = [\mathbf{A} \;\; \mathbf{AB}^* \;\; \mathbf{AX}]$. Thus, (A-1) becomes

$$\mathbf{M} = [\mathbf{A} \;\; \mathbf{AB}^* \;\; \mathbf{AX}] \left(\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{m}^*\mathbf{m}^{*\prime} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{vv}' \end{bmatrix} + \begin{bmatrix} \mathbf{G}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \right) \begin{bmatrix} \mathbf{A}' \\ \mathbf{B}^{*\prime}\mathbf{A}' \\ \mathbf{X}'\mathbf{A}' \end{bmatrix} + \mathbf{K}^2$$

$$= \mathbf{A}(\mathbf{G}^2 + \mathbf{B}^*(\mathbf{m}^*\mathbf{m}^{*\prime} + \mathbf{D}^2)\mathbf{B}^{*\prime} + \mathbf{X}(\mathbf{vv}' + \mathbf{I})\mathbf{X}')\mathbf{A}' + \mathbf{K}^2,$$

which is identical to the moment structure required by (24).

The above specification is apparently not unique. For example, setting

$$\Lambda_y = [\mathbf{A} \;\; \mathbf{0} \;\; \mathbf{0}]$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{B}^* & \mathbf{X} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

will give the same result. This latter specification may be more general than Result 2 in that it does not assume that both **A** and $\mathbf{B}^*$ are known *a priori*. However, in Result 3 both **A** and $\mathbf{B}^*$ have to be assumed known *a priori*.

Result 3. The moment structure of $\tilde{\mathbf{y}}$ (**t** in our notation) corresponding to (25) is given by setting

$$\Lambda_y = [\mathbf{A} \;\; \mathbf{AB}^* \;\; \mathbf{0}]$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\Gamma = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{v} \end{pmatrix}$$

and

$$\Psi = \begin{bmatrix} \mathbf{G}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

(Proof) $\Lambda_y(\mathbf{I} - \mathbf{B})^{-1} = [\mathbf{A} \;\; \mathbf{AB}^* \;\; \mathbf{AB}^*\mathbf{Q}]$. Thus, (A-1) becomes

$$\mathbf{M} = [\mathbf{A} \;\; \mathbf{AB}^* \;\; \mathbf{AB}^*\mathbf{Q}] \left(\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{vv}' \end{bmatrix} + \begin{bmatrix} \mathbf{G}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \right) \begin{bmatrix} \mathbf{A}' \\ \mathbf{B}^*\mathbf{A}' \\ \mathbf{Q}'\mathbf{B}^*\mathbf{A}' \end{bmatrix} + \mathbf{K}^2$$

$$= \mathbf{A}(\mathbf{G}^2 + \mathbf{B}^*\mathbf{Q}(\mathbf{vv}' + \mathbf{I})\mathbf{Q}'\mathbf{B}^{*\prime})\mathbf{A}' + \mathbf{K}^2,$$

which is identical to the moment structure required of (25). A slight generalization can be made by setting

$$\Psi = \begin{bmatrix} \mathbf{G}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

The moment structure then becomes

$$\mathbf{M} = \mathbf{A}(\mathbf{G}^2 + \mathbf{B}^*(\mathbf{Q}(\mathbf{vv}' + \mathbf{I})\mathbf{Q}' + \mathbf{D}^2)\mathbf{B}^{*\prime})\mathbf{A}' + \mathbf{K}^2.$$

References

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.

Bechtel, G. G., Tucker, L. R., & Chang, W. (1971). A scalar products model for the multidimensional scaling of choice. *Psychometrika, 36,* 369-387.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika, 48,* 493-517.

Bentler, P. M. (1985). *Theory and implementation of EQS, a structural equations program.* Los Angeles: BMDP Statistical Software.

Bloxom, B. (1972). The simplex in pair comparisons. *Psychometrika, 37,* 119-136.

Bock, R. D., & Bargman, R. E. (1966). Analysis of covariance structures.

Psychometrika, 31, 507-534.

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

Bradley, R. A. (1976). Science, statistics and paired comparisons. *Biometrics, 32,* 213-232.

Bradley, R. A., & Terry, M. E. (1952). The rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika, 39,* 324-345.

Browne, M. W. (1974). Generalized least squares estimates in the analysis of covariance structures. *South African Statistical Journal, 8,* 1-24.

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: Hans Huber.

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

Debreu, G. (1960). Review of R. D. Luce, Individual choice behavior: A theoretical analysis. *American Economic Review, 50,* 186-188.

De Soete, G. (1983). On the relation between two generalized cases of Thurstone's law of comparative judgment. *Mathématiques et Sciences humaines, 21,* 45-57.

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 comparisons data. *Journal of Mathematical Psychology, 30,* 28-41.

Guttman, L. (1946). An approach for quantifying paired comparisons rank order. *Annals of Mathematical Statistics, 17,* 144-163.

Halff, H. M. (1976). Choice theories for differentially comparable alternatives. *Journal of Mathematical Psychology, 14,* 244-246.

Heiser, W., & De Leeuw, J. (1981). Multidimensional mapping of preference data. *Mathématiques et Sciences humaines, 19,* 39-96.

Hohle, R. H. (1966). An empirical evaluation and comparison of two models for discriminability. *Journal of Mathemathical Psychology, 3,* 174-183.

Indow, T. (1975). On choice probability. *Behaviormetrika, 2,* 13-31.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika, 57,* 239-251.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409-426.

Jöreskog, K. G., & Sorbom, D. (1981). *LISREL VI user guide.* Mooresville, IN: Scientific Software.

Krantz, D. H. (1967). Small-step and large-step color differences for monochromatic stimuli of constant brightness. *Journal of the Optical Society of America, 57,* 1304-1316.

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

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology, 15,* 215-233.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology, 37,* 234-251.

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.

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

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

Muthen, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46,* 407-419.

Restle, F. (1961). *Psychology of judgment and choice.* New York: Wiley.

Rumelhart, D. L., & Greeno, J. G. (1971). Similarity between stimuli: An experimental test of the Luce and Restle choice models. *Journal of Mathematical Psychology, 8,* 370-381.

Sjöberg, L. (1975). Uncertainty of comparative judgments and multidimensional structure. *Multivariate Behavioral Research, 11,* 207-218.

Sjöberg, L. (1977). Choice frequency and similarity. *Scandinavian Journal of Psychology, 18,* 103-115.

Sjöberg, L. (1980). Similarity and correlation. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice.* Bern: Hans Huber.

Sorbom, D. (1981). Structural equation models with structured means. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observations: Causality, structure and prediction.* Amsterdam: North-Holland.

Strauss, D. (1981). Choice by features: An extension of Luce's model to account for similarities. *British Journal of Mathematical and Statistical Psychology, 34,* 50-61.

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. (1985). Probabilistic multidimensional pair comparison models that take into account systematic individual differences. Transcript of the talk given at the 50th Anniversary Meeting of the Psychometric Society, Nashville.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273-286.

Thurstone, L. L. (1959). *The measurement of values.* Chicago, IL: University of Chicago Press.

Tversky, A. (1972a). Choice by elimination. *Journal of Mathematical Psychology, 9,* 341-367.

Tversky, A. (1972b). Elimination by aspects: A theory of choice. *Psychological Review, 79,* 281-299.

Tversky, A., & Russo, J. E. (1969). Substitutability and similarity in binary choices. *Journal of Mathematical Psychology, 6,* 1-12.

Tversky, A., & Sattath, S. (1979). Preference trees. *Psychological Review, 86,* 542-573.

Yanai, H. (1970). Factor analysis with external criteria. *Japanese Psychological Research, 12,* 143-153.