THE WEIGHTED ADDITIVE MODEL

Yoshio Takane

Reprinted from Law, H.G. et al. (Eds.)
Research methods for multimode data analysls.
New York: Praeger, 1984.

13

The Weighted Additive Model

Yoshio Takane

There are a number of situations in which various attributes of stimuli are to be weighted differently for different decision purposes. For example, consider a person buying a car. Desirable properties of a car depend on many factors. Spaciousness may be an important factor for comfortable weekend driving, while cost efficiency—such as gas mileage—may be more important for commuting. In any case, the person has to decide what is optimal for him, considering the various situational demands surrounding him. One way to achieve this is to express his best interest as a linear combination of various attributes of cars. Then the cars can be evaluated and compared on the composite criterion. The weights to be given to the attributes, of course, depend on his specific interest.

As another example, suppose a psychology department would like to set up different admission criteria for different subprograms it offers. Again, this may be done by forming a composite criterion most suitable for each program, taking into account various aspects of each applicant's qualifications. For instance, if the applicant is seeking admission in quantitative psychology, his score on GRE-Quantitative should be weighed heavily in forming the composite score. If, on the other hand, the prospective student is going into clinical psychology, mathematical skill is perhaps not as important; consequently, it should be weighed much less. The problem of weighting various attributes of stimuli into composite scores is called a *weighting problem*. The problem is called a *differential weighting problem*, when two or more composite scores are derived from the same set of attribute scores according to multiple criteria. Assuming that the composition rule is linear, the weighting problem may expressed as:

Comments by Tony Marley have been helpful in revising this manuscript.

$$y_{kr} = \sum_{t=1}^{T} w_{kt} x_{rt} , \qquad (k = 1, \ldots, K) , \qquad (13\text{–}1)$$

where y_{kr} is the composite score of stimulus r for criterion k, x_{rt} is the score of stimulus r on attribute t (assumed known a priori), and w_{kt} is the weight attached to attribute t for criterion k. The weights are determined in such a way that the derived composite scores (y_{kr}) best agree with the criterion variables. If the criterion variables can be assumed linearly related to the attribute scores, the weights in (13–1) can be determined by applying the linear regression analysis separately to each criterion (k). Thus, there is no need for a new analysis method in this case.

Note that in (13–1) attribute score x_{rt} is assumed known a priori. Most often it represents a numerical value designating a degree to which stimulus r possesses attribute t. However, there are cases in which attributes are not quantified a priori. For example, types of cars—such as sedan or wagon—may be an important factor in deciding which car to buy. But there is no natural and obvious way of assigning a numerical value to the attribute. It should be quantified before it can be used in a composite criterion. An attribute may sometimes have to be requantified, even when a priori quantifications exist. For example, performance on GRE-Quantitative is usually reported in a numerical form, but there is no assurance that the reported score is linearly related to criterion variables, such as performance in quantitative psychology. However, the linear form of (13–1) implies that x_{rt} is linearly related to the criterion variables. Thus, it is necessary to requantify x_{rt} in such a way that it is as linearly related to the criterion variables as possible.

The quantifications of an attribute may generally be expressed as follows. Let h_{rq_t} denote an indicator variable such that:

$$h_{rq_t} = \begin{cases} 1, \text{ if stimulus } r \text{ belongs to level } q \text{ of attribute } t \\ \quad \text{(this may be subject } r \text{ passing item } q \text{ in test } t.); \\ 0, \text{ otherwise.} \end{cases}$$

Then the new attribute score x_{rt} is obtained by a linear combination of the indicator variables; that is,

$$x_{rt} = \sum_{q_t}^{Q_t} h_{rq_t} \alpha_{q_t} , \qquad (13\text{–}2)$$

where α_{q_t} is the score given to the qth level of attribute t, and Q_t is the total number of levels in the attribute. Specifically, we are discretizing the attribute into several levels and assigning new numerical values at these discrete points of the attribute.

If we substitute the expression (13–2) for x_{rt} in (13–1), we obtain:

$$y_{kr} = \sum_{t}^{T} w_{kt} \left(\sum_{q_t}^{Q_t} h_{rq_t} \alpha_{q_t} \right) , \quad (k = 1, \ldots, K) , \qquad (13\text{–}3)$$

This model is called the *weighted additive model* (WAM) (Takane, Young, and de Leeuw 1980). In (13–3), the differential weighting problem (13–1) and the quantification problem (13–2) (which is yet another weighting problem) are combined into one major problem, in which the two subproblems are solved simultaneously. In (13–3), both α_{q_t} and w_{kt} are estimated so that the derived composite scores, y_{kr} $(k = 1, \ldots, K)$, are best predictive of all K criteria simultaneously.

When there is only one criterion—that is, $K = 1$—we obtain:

$$y_r = \sum_{t}^{T} \sum_{q_t}^{Q_t} h_{rq_t} \alpha^*_{q_t} , \qquad (13\text{–}4)$$

where $\alpha^*_{q_t} = \alpha_{q_t} w_t$. Notice that subscript k has been omitted.

This model is called the *simple additive model* (SAM). In this case, we have only the quantification problem, which reduces to a linear regression problem with $h_{rq_t}(q = 1, \ldots, Q)$ serving as the independent variables.

Note that in (13–3), α_{qt} is assumed common across all criteria, just as in (13–1) the same x_{rt} is used for all criteria. If, on the other hand, α_{q_t} is allowed to vary over the criteria, we have:

$$y_{kr} = \sum_{t}^{T} w_{kt} \left(\sum_{q_t}^{Q_t} h_{rq_t} \alpha_{kq_t} \right) \qquad (13\text{–}5)$$

$$= \sum_{t}^{T} \sum_{q_t}^{Q_t} h_{rq_t} (\alpha_{kq_t} w_{kt}) = \sum_{t} \sum_{q_t} h_{rq_t} \alpha^*_{kq_t},$$

where $\alpha^*_{kq_t} = \alpha_{kq_t} w_{kt}$. Note that subscript k is attached to α_{kq_t} and $\alpha^*_{kq_t}$, implying that they are unique to each k. Equation (13–5) indicates that the WAM in this case reduces to the SAM applied separately to each criterion. The new quantification parameter, $\alpha^*_{kq_t}$, can be estimated by applying the linear regression analysis separately to each k. (This is completely analogous to [13–1], in which the weights are obtained by separate regression analyses of y_{kr} to x_{rt}.) This means that a real significant feature of the WAM lies in the assumption of common α_{q_t} across different criteria. This very feature calls for a new analysis method specifically designed for this model, since no conventional multivariate methods apply to this case. In this chapter, we discuss various aspects of the WAM focusing on its empirical motivation, mathematical properties, analytical methods, and examples of application.

EMPIRICAL MOTIVATION

As has been emphasized, an important assumption underlying the WAM is that α_{q_t} remains constant across different criteria.

Although this assumption is implicit in almost all the models that capitalize on a priori quantifications of attributes, it by no means is always correct. Rather, it is an empirical hypothesis to be verified or falsified in each specific situation in which the model is employed. On the other hand, there are many situations in which the WAM is indeed appropriate. In this section, we discuss a couple more examples of such situations. Several more examples will be given in the application section.

The first example has been modified from Kruskal and Shepard (1974). Suppose first that a set of cylinders are generated by combining different altitude levels and base areas. Suppose further that those cylinders are measured in terms of two physical variables, side area (y_S) and volume (y_V). Let α and β represent altitude and base area, respectively. Then we have:

$$y_S = (2\sqrt{\pi})\,\alpha\,\beta^{1/2} ,$$

and

$$y_V = \alpha\,\beta .$$

By the log transformation, we obtain:

$$\log(y_S) = \log(\alpha) + \tfrac{1}{2}\log(\beta) + \gamma , \qquad (13\text{–}6)$$

and

$$\log(y_V) = \log(\alpha) + \log(\beta) , \qquad (13\text{–}7)$$

where $\gamma = \log(2\sqrt{\pi})$. That is, the log-transformed y_S and y_V are both additive functions of log (α) and log (β), disregarding γ, for the moment. The difference between the two criteria—log (y_S) and log (y_V)—is produced by a differential weighting of log (α) and log (β), which themselves remain constant across the criteria. It can be readily seen that (13–6) and (13–7) are special cases of (13–1).

Suppose now that we are ignorant of the exact weighting rules to obtain the two criteria—log (y_S) and log (y_V)—from log (α) and log (β), except that the former are both additive functions of the latter. The problem of finding the differential weights, given a set of measurements on the four variables, is identical to the one posed by (13–1). Now assume further that actual numerical values of log (α) and log (β) are not directly observable, but only discrete levels of altitude and base area of the cylinders are provided as information. Then, those levels have to be quantified before they can be used for log (α) and log (β) in (13–6) and (13–7). The quantifications of the levels should stay invariant, just as both log (α) and log (β) remain intact in (13–6) and (13–7). The problem now is equivalent to what we have in (13–3): Namely, we simultaneously quantify the levels of the attribute variables (altitude and base area) and estimate the differential weights to obtain log (y_S) and log (y_V).

Finally, let us suppose that y_S and y_V, rather than log (y_S) and log (y_V), are observed as the criterion variables. The observed criteria cannot be represented by any additive functions

of log (α) and log (β) without transformations. But suppose we are ignorant of the transformations that make y_S and y_V additively representable by log (α) and log (β). Then we are faced with an additional problem of finding the transformations. The transformations may be found by quantifying or requantifying the criterion variables (possibly with ordinal restrictions) in much the same way as the attribute variables are quantified.

As a second example, suppose the desirability (u) of various cars is investigated as a function of their price (p) and fuel efficiency (g). Although both price and fuel efficiency can be measured physically, their contributions to the overall desirability ratings may not be linear. Thus, several price levels and fuel efficiency levels are chosen, from which a set of hypothetical cars (such as a car costing $7,000, featuring high fuel efficiency, and so on) are constructed for the desirability ratings. Let $f_P(p_i)$ and $f_G(g_j)$ represent quantifications of the ith price level (p_i) and the jth fuel efficiency level (g_j), respectively. It is likely that f_P is monotonically decreasing with price and f_G monotonically increasing with fuel efficiency, although neither of them may be linear with their physical correlates. Let the observed rating of a car characterized by p_i and g_j be denoted by u_{ij}. Let us assume that u_{ij} is monotonically related to the sum of $f_P(p_i)$ and $f_G(g_j)$. That is,

$$f_U(u_{ij}) = f_P(p_i) + f_G(g_j) \ , \tag{13–8}$$

where f_U is some monotonic function. (We have tentatively disregarded possible measurement errors in [13–8].) This model implies that after the desirability rating (the criterion variable) is monotonically transformed by an appropriate transformation, it can be represented as an additive function of f_P and f_G (the attribute variables). Model (13–8) is, in fact, a SAM.

Suppose now that several individuals have rated the desirability of the same set of cars, and a substantial amount of individual differences have been observed. The differences may be due to possible differences in the attribute functions; that is, the way price levels and fuel efficiency levels are evaluated may vary across individuals. In this case, model (13–8) should be applied separately to each individual. On the other hand, the differences may only be due to possible differences in relative importance of the two attributes, while the evaluations of the attributes themselves remain more or less constant across the individuals. If this is the case, we obtain:

$$f_{kU}(u_{ij}) = w_{kP} f_P(p_i) + w_{kG} f_G(g_j) \ , \tag{13–9}$$

where k refers to the kth individual. The differences among the individuals are accounted for by a differential weighting of the attribute variables, f_P and f_G, which are additive and invariant across k. Model (13–9) is the WAM. (Note that individual differences may be allowed in f_U, as indicated by subscript k attached.)

The key question, then, is whether or not f_P and f_G remain constant across different individuals. On an intuitive basis this seems likely. The ratings concerned with desirability (and not

preference) and the desirability perceptions are much more apt to agree. For example, nobody would disagree that f_P is monotonically decreasing and f_G is monotonically increasing, although ultimately this is a question to be answered on an empirical basis. Whether it is more appropriate to apply the SAM separately to each individual or to apply the WAM jointly to all individuals can only be answered by fitting the two models to the same set of data and by comparing their goodness-of-fit to the data. If the fit of the SAM is significantly better than that of the WAM, it is not appropriate to assume f_P and f_G are constant across k. Otherwise, the WAM is more appropriate, because it uses fewer parameters than the SAM fitted separately.

THE MODEL

We now turn to investigating various properties of the WAM. Our discussion concentrates on formal aspects of the model, but it may be helpful for the reader to bear in mind some concrete example to which the model might be applied.

Basic Properties

We first restate the model in standard notation. For explanatory convenience, we state the model for the simple two-factor case. We also introduce new terminologies.

Let α_i and β_j denote quantifications of the ith level of attribute A and the jth level of attribute B, respectively. They are sometimes called *additive effects*. (Attributes are often referred to as *factors* or *independent variables*.) Let w_{kA} and w_{kB} represent the weights attached to factors A and B, respectively, for the kth criterion variable. (The criterion variable is often called the *dependent variable*.) Then the WAM—which has already been stated in such forms as (13–3) and (13–9)—can be restated as:

$$y_{kij} = w_{kA}\alpha_i + w_{kB}\beta_j \,, \qquad (k = 1, \ldots, K) \,, \quad (13\text{–}10)$$

where y_{kij} is the model prediction or model value representing the combined effect of the ith level of factor A (α_i) and the jth level of factor B (β_j) in the kth criterion variable. We assume that there are n_A levels in factor A ($i = 1, \ldots, n_A$) and n_B levels in factor B ($j = 1, \ldots, n_B$). An extension of (13–10) to a higher-order design is straightforward. What have been referred to as the *criteria* or *criterion variables* may represent whatever sets of observations that need be distinguished and whose distinction is best characterized by a differential weighting. For example, they may represent subjects, groups, occasions, experimental conditions, and so on. There is an intrinsic scale indeterminacy in the WAM. That is, for each factor, say A, the effect of multiplying w_{kA} by some constant c ($\neq 0$) is offset by dividing α_i by the same constant. We may remove this indeterminacy by requiring

$$\sum_{i=1}^{n_A} \alpha_i^2 = n_A \quad \text{and} \quad \sum_{j=1}^{n_B} \beta_j^2 = n_B \ . \tag{13–11}$$

Also, the additive effects can be determined only up to an arbitrary origin. We thus require:

$$\sum_{i=1}^{n_A} \alpha_i = 0 \quad \text{and} \quad \sum_{j=1}^{n_B} \beta_j = 0 \ . \tag{13–12}$$

Consequently,

$$\sum_i \sum_j y_{kij} = 0 \qquad \text{for all } k.$$

This amounts to centering the model predictions so that their mean is equal to zero.

To reemphasize, the WAM ([13–10]) postulates that the differences among K sets of observations arise from a differential weighting of additive effects that remain constant across the sets. Reflecting the fact that the WAM was originally introduced for explaining individual differences in perception (see the application section), the differential weights, w_{kA} and w_{kB}, are often called *individual differences weights*. When there are no systematic differences among the sets, the individual differences weights can be set to unity, and (13–10) reduces to the SAM,

$$y_{kij} = \alpha_{ki} + \beta_{kj} , \tag{13–13}$$

which treats k as a replication factor. At the other extreme, the additive effects may be completely unique for each set, in which case each set must be represented separately by the SAM,

$$y_{kij} = \alpha_{ki} + \beta_{kj} , \qquad (k = 1, \ . \ . \ . \ , K) , \tag{13–14}$$

where α_{ki} and β_{kj} represent the additive effects unique to set k. The WAM lies somewhere between these two extreme cases; it represents both the common and unique aspects of multiple conjoint data in one model. (The data supposedly representing the combined effects of two or more influencing factors are sometimes called *conjoint data* after Shepard [1972].)

We now explore some basic properties of the WAM. We have already mentioned that in order for the WAM to apply to K sets of observations, each set has to be represented by the SAM. Suppose that we have applied the SAM separately to each data set and obtained estimates of the additive effects, α_{ki} and β_{kj}, $(k = 1, \ . \ . \ . \ , K)$. From these, can we tell whether the WAM is appropriate for all sets of observations?

Since the WAM is a special case of the SAM with $\alpha_{ki} = w_{kA}\alpha_i$ $(i = 1, \ . \ . \ . \ , n_A)$ and $\beta_{kj} = w_{kB}\beta_j$ $(j = 1, \ . \ . \ . \ , n_B)$ the additive effects from the SAM have to be proportional; namely,

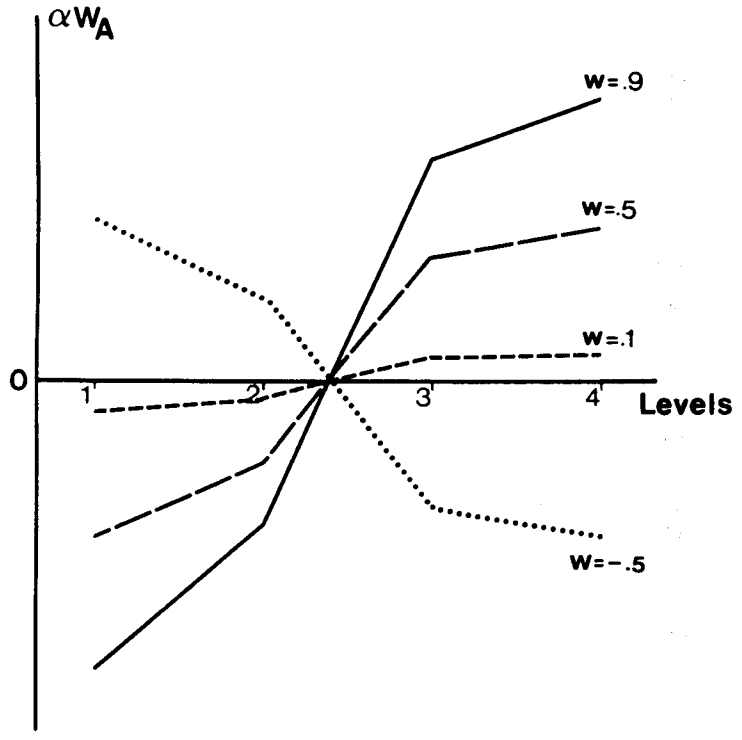
$$\frac{\alpha_{kj}}{\alpha_{k'j}} = \frac{w_{kA}}{w_{k'A}} \qquad \text{for all } j, \tag{13–15}$$

and

$$\frac{\beta_{kj}}{\beta_{k'j}} = \frac{w_{kB}}{w_{k'B}} \qquad \text{for all } j.$$

The situation is depicted in Figure 13–1, where $\alpha_{kj} = w_{kA}\alpha_j$ ($j = 1, \ldots, n_A$) are plotted for different values of w_{kA} and connected for each k by line segments. The plots exhibit some characteristic pattern. Note that the additive effects were tacitly assumed monotonically increasing in Figure 13–1, but this does not imply that it is always the case. For the monotonically increasing α_j, α_{kj} is monotonically increasing for a positive w_{kA} and monotonically decreasing for a negative w_{kA}. The rate of increase and decrease changes as a function of the absolute value of w_{kA}. (Of course, the change rate is higher for a larger value of w_{kA}

Figure 13–1. The Plot of $\alpha_{kj} = \alpha_j w_{kA}$ ($j = 1, \ldots, n_A$) when the WAM is Correct



and vice versa.) All the curves in Figure 13–1 have essentially the same profile (shape) except that their scatters (Cronbach and Gleser 1953) are different. In fact, $|w_{kA}|$ may be interpreted as variance of α_{kj} over j. All the curves cross with one another whenever they cross a zero line. (They should cross the zero line at least once to meet restriction [13–12].) Note that Figure 13–1 was drawn for an ideal case in which the WAM fits the data perfectly. In practical situations, it is not possible to observe error-free data. Consequently, the proportionalities of α_{kj} and β_{kj} may hold only approximately.

Some basic properties of the WAM may be directly observed in the original data themselves, even without estimating α_{kj} and β_{kj}. In order to see this, let us assume some hypothetical parameter values and see if there is any characteristic pattern appearing in the model predictions. For simplicity, we assume $n_A = n_B = 3$; $K = 4$; $\alpha_1 = 1$, $\alpha_2 = 3$, and $\alpha_3 = 4$; and $\beta_1 = 2$, $\beta_2 = 3$, and $\beta_3 = 5$. The four sets of model predictions derived from the model are given in Figure 13–2. Note that the hypothesized values of the additive effects do not satisfy (13–11) or (13–12), but this should not matter for the purpose of examining properties of the WAM. These restrictions are merely conventional and the basic properties of the model remain unaffected even if they are violated. Note also that both α_j and β_j are monotonically increasing, but again, this is mere coincidence. In Figure 13–2, $w_{kA}\alpha_j = \alpha_{kj}$ and $w_{kB}\beta_j = \beta_{kj}$ are entered in row and column marginals of each table, respectively. Main entries in the table are the model predictions (y_{kij}) calculated by (13–10).

Figure 13–2. Examples of Predictions from the Weighted Additive Model (WAM)

$$\alpha = \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}, \quad \beta = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}$$

	w_A	w_B
k=1	2	1
k=2	1	2
k=3	-1	-1
k=4	1	-2

k=1

$w_{1B}\beta$ \ $w_{1A}\alpha$	2 <	6 <	8
2	4 <	8 <	10
3	5 <	9 <	11
4	7 <	11 <	13

k=2

$w_{2B}\beta$ \ $w_{2A}\alpha$	1 <	3 <	4
4	5 <	7 <	8
6	7 <	9 <	10
10	11 <	13 <	14

k=3

$w_{3B}\beta$ \ $w_{3A}\alpha$	-1 >	-3 >	-4
-2	-3 >	-5 >	-6
-3	-4 >	-6 >	-7
-5	-6 >	-8 >	-9

k=4

$w_{4B}\beta$ \ $w_{4A}\alpha$	1 <	3 <	4
-4	-3 <	-1 <	0
-6	-5 <	-3 <	-2
-10	-9 <	-7 <	-6

Since the WAM is a special case of the SAM for each data set, all the properties satisfied by the WAM must be satisfied by the SAM for each k. Let us, therefore, first see what characteristic patterns may be observed for each data set. They should be the patterns we expect to observe for each data set to be represented by the SAM.

Let us start with a quantitative relation among y_{kij}. It can be immediately observed that:

$$y_{kij} = \bar{y}_{ki.} + \bar{y}_{k.j} - \bar{y}_{k..} \qquad \text{for all } i \text{ and } j,$$

where

$$\bar{y}_{ki.} = \sum_{j}^{n_B} y_{kij}/n_B ,$$

$$\bar{y}_{k.j} = \sum_{i}^{n_A} y_{kij}/n_A ,$$

and

$$\bar{y}_{k..} = \sum_{i}^{n_A} y_{ki.}/n_A = \sum_{j}^{n_B} y_{k.j}/n_B = \sum_{i}^{n_A}\sum_{j}^{n_B} y_{kij}/n_A n_B .$$

Not surprisingly, this is equivalent to the noninteraction condition in two-way ANOVA. This condition states that the combined effect (y_{kij}) of the ith level of factor A and the jth level of factor B is a simple addition of a contribution from the ith level of factor A ($\bar{y}_{ki.}$) and a contribution from the jth level of factor B ($\bar{y}_{k.j}$). ($\bar{y}_{k..} = 0$ if α_i and β_j have satisfied [13–12].)

If the observations are linearly related to the model predictions (that is, if they are obtained on an interval scale), the above condition should still be true approximately, if not exactly. However, it does not provide a valid criterion for an additive representation if the data are only monotonically related to the model predictions.

Are there any patterns in the table that are invariant over any monotonic distortion of the model predictions? Observe that for each k, $w_{kA}\alpha_i$ is either ascending or descending, depending on the sign of w_{kA}. The same is true for $w_{kB}\beta_j$. What is more important is that the model predictions are in the same order as $w_{kA}\alpha_i$ within each row, and they are in the same order as $w_{kB}\beta_j$ within each column. That is, if

$$y_{kij} \geq y_{kij'} ,$$

then

$$y_{ki'j} \geq y_{ki'j'} \qquad \text{for all } i'.$$

And if

$$y_{kij} \geq y_{ki'j} ,$$

then

$$y_{kij'} \geq y_{ki'j'} \qquad \text{for all } j',$$

for each k.

In the additive conjoint measurement literature (Luce and Tukey 1964; Krantz et al. 1971), the above condition applied to a set of observations is called the *independence* or *single cancellation condition*. It is one of necessary conditions for an additive representation of ordinal data.

Another necessary condition, called the *Thomsen condition*, may not be obvious in the table. This condition states that if

$$y_{kij} = y_{kpq} \quad \text{and} \quad y_{kpr} = y_{ksj},$$

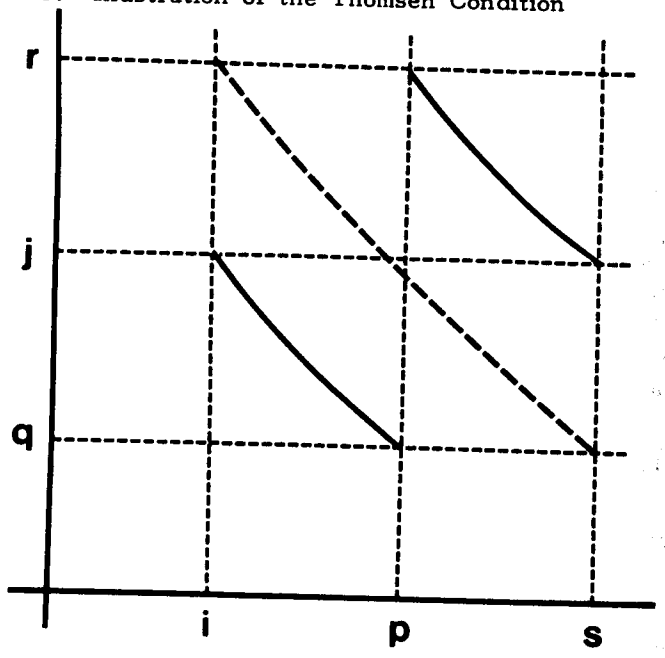
then

$$y_{kir} = y_{ksq},$$

in terms of the model predictions. This is illustrated in Figure 13–3, where indifference relations in the premise are connected by solid line segments and the indifference relation in the conclusion by a dotted line segment. Unfortunately, it is difficult to verify this condition directly, since we may not be able to find stimulus pairs—$(i,j),(p,q)$ and $(p,r),(s,j)$—such that they satisfy the indifference relations postulated in the premise.

The independence and the Thomsen conditions, along with other conditions not described here (including weak-ordering,

Figure 13–3. Illustration of the Thomsen Condition



essentialness, solvability, and so on) ensure a representation of ordinal data by the SAM. Those are the qualitative properties that have to be satisfied in order to be representable by the SAM. Since actual data usually contain errors, it may not be easy to determine if these properties are met to a satisfactory degree. Nevertheless, they should serve as rough guidelines.

The additive representation in turn implies the double-cancellation condition, which is just the Thomsen condition with "=" replaced by "≥." (The Thomsen condition is actually a special case of the double-cancellation condition; both "≥" and "≤" imply "=".) The double-cancellation condition can be directly tested. We see that it is satisfied by the model predictions, y_{kij}. However, even though the double-cancellation condition is satisfied by a particular data set, this does not imply that the Thomsen condition is satisfied, although the former is stronger than the latter. This is because the data are always finite and a particular condition being satisfied by *available* data points does not ensure that it is satisfied by *all conceivable* data points. This means that we may falsify a condition by observing an instance of its violation, but we can never completely verify the condition based on a finite sample of observations. (This is similar to the statistical hypothesis testing situation in which one can never say the null hypothesis is true.)

The WAM requires more than is required by the SAM. Let us again begin with a quantitative property and then relax it into ordinal properties. We observe that:

$$(\bar{y}_{ki.} - \bar{y}_{k..})/(\bar{y}_{k'i.} - \bar{y}_{k'..}) = \text{constant for all } i,$$

and

$$(\bar{y}_{k.j} - \bar{y}_{k..})/(\bar{y}_{k'.j} - \bar{y}_{k'..}) = \text{constant for all } j,$$

which are analogous to (13–15). Since a column or row mean minus a grand mean is an estimate of α_{ki} (or β_{kj}), the proportionalities of the additive effects obtained from the SAM imply the above condition. (Again, $\bar{y}_{k..} = \bar{y}_{k'..} = 0$ if [13–12] were satisfied. In this case, the above condition would be much simplified.)

The condition should be relaxed for ordinal data. Looking at Figure 13–2 across tables, we may observe that the ordering among the model predictions within each row or column is either exactly the same or opposite across columns or rows, depending on whether the signs of the weights are in agreement or in disagreement. For example, $4 < 8 < 10$ in the first row of $k = 1$, $5 < 7 < 8$ in the first row of $k = 2$ (same order), and $-3 > -5 > -7$ in the first row of $k = 3$ (reverse order). Thus, it may be stated that if

$$y_{kij} \geq y_{kij'},$$

then either

$$y_{k'ij} \geq y_{k'ij'} \quad \text{or} \quad y_{k'ij} \leq y_{k'ij'}.$$

And if

$$y_{kij} \geq y_{ki'j} ,$$

then either

$$y_{k'ij} \geq y_{k'i'j} \quad \text{or} \quad y_{k'ij} \leq y_{k'i'j} .$$

The above condition, stated in terms of observed data (instead of model predictions), is called *homogeneous ordering* (Sayeki 1972), and is one of the necessary properties of the WAM. This condition is analogous to the independence condition for the SAM.

Another necessary condition, which parallels the Thomsen condition for the SAM, is a version of the *Reidemeister condition* (Krantz et al. 1971), a special case of the triple-cancellation axiom. This condition states that if

$$y_{kij} = y_{kpq} ,$$

$$y_{kir} = y_{kpt} ,$$

and

$$y_{k'sr} = y_{k'ut} ,$$

then

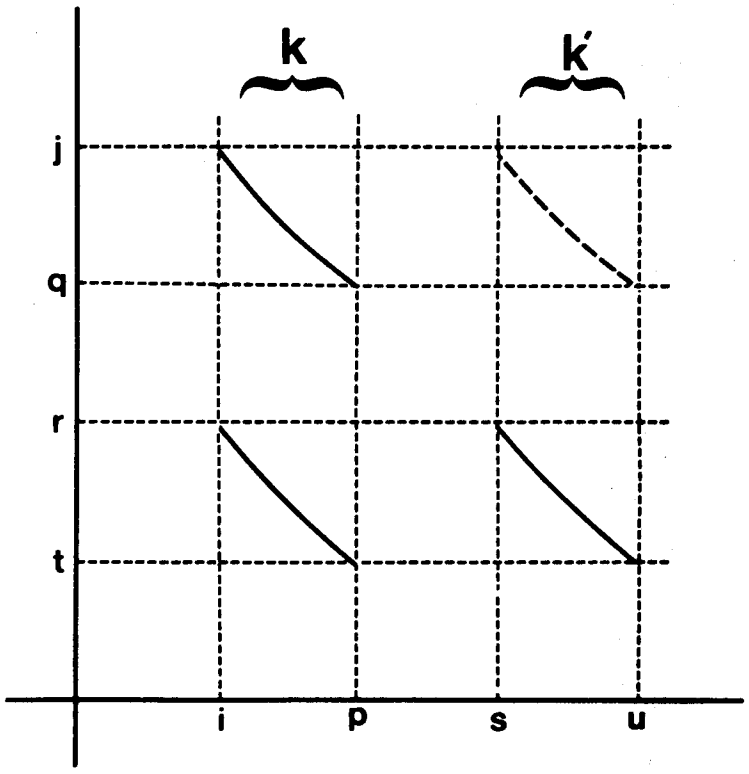
$$y_{k'sj} = y_{k'uq} ,$$

as depicted in Figure 13–4. Again, indifferences in the premise are connected by a solid line and the indifference in the conclusion is connected by a dotted line. For the same reason that the Thomsen condition was difficult to verify, this condition may not be testable in practical situations. However, it has been proven (Sayeki 1972; Marley 1970) that the homogeneous ordering and the Reidemeister conditions together with other structural axioms ensure an existence of a representation of ordinal data by the WAM. Just as the SAM implied the double-cancellation condition, the WAM implies the triple-cancellation condition (which is the Reidemeister condition with "=" replaced by "≥"). This condition, as well as the homogeneous ordering condition, can be directly tested with observed data.

Geometric Properties

When n_A levels of factor A and n_B levels of factor B are factorially combined to construct a set of stimuli, we obtain $n_A \times n_B$ stimuli altogether and consequently as many model predictions for each k. These $n_A n_B$ model predictions can be put into a vector form according to a certain prescribed order. This vector may be denoted by $\mathbf{y}_k$ $(k = 1, \ldots, K)$. Let $G_A(n_A n_B \times n_A)$ and $G_B(n_A n_B \times n_B)$ denote design matrices for factors A and B, respectively. That is, the rth row of G_A has one in the ith column and zeroes elsewhere if the rth element of $\mathbf{y}_k$ (y_{kr}) is related to

Figure 13–4. Illustration of a Triple-Cancellation Condition

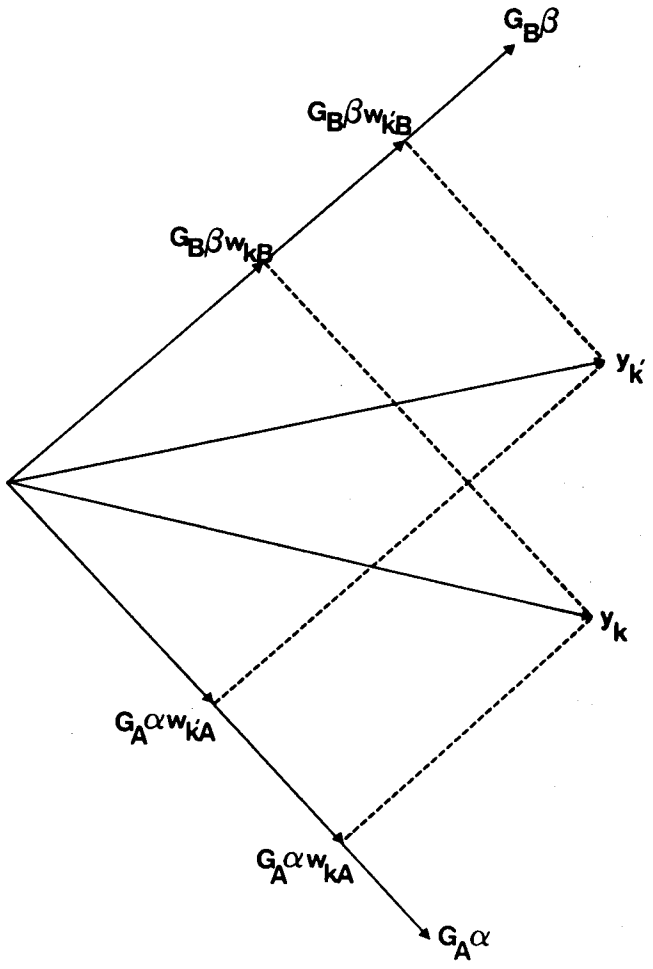


the ith level of factor A. Similarly, the rth row of G_B has one in the jth column and zeroes elsewhere if y_{kr} comes from the jth level of factor B. Let α and β be vectors of the additive effects (α_i and β_j) for factors A and B, respectively. Then the entire set of model predictions can be expressed as:

$$\mathbf{y}_k = G_A\alpha w_{kA} + G_B\beta w_{kB} \ , \quad (k = 1, \ldots, K) \ . \qquad (13\text{–}16)$$

The above equation indicates that $\mathbf{y}_k$ is the weighted sum of two vectors, $G_A\alpha$ and $G_B\beta$, the former being weighted by w_{kA} and the latter by w_{kB} ($k = 1, \ldots, K$). The two constituent vectors remain the same across k, while the weights applied to the vectors change over k. This implies, geometrically, that all $\mathbf{y}_k$'s ($k = 1, \ldots, K$) lie in the space spanned by the two constituent vectors, $G_A\alpha$ and $G_B\beta$ (which themselves are linear combinations of G_A and G_B, respectively), their directions being subject to change as functions of the weights. The situation is depicted in Figure 13–5 for the case of two criteria (k and k'). In order to obtain $\mathbf{y}_k$, $G_A\alpha$ is shrunk or stretched by w_{kA} and $G_B\beta$ by w_{kB} before being added together. Similarly, $G_A\alpha$ is shrunk by $w_{k'A}$ and $G_B\beta$ by $w_{k'B}$ to obtain $\mathbf{y}_{k'}$. This explains the mechanism that generates two or more distinct vectors of model predictions from

Figure 13–5. Geometric Interpretation of the WAM



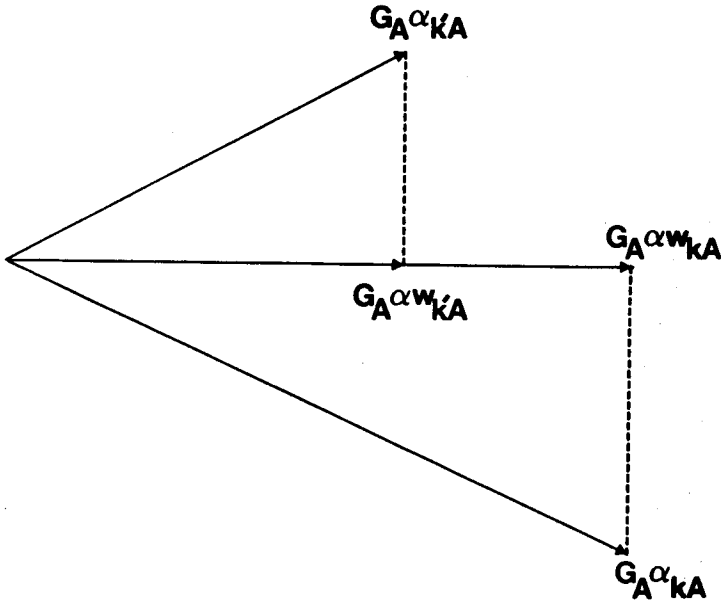
the same attribute vectors in the WAM.

The geometrical relationship between the WAM and the SAM may be explored in a similar manner. For this, the SAM may be written in a vector form analogous to (13–16):

$$y_k = G_A\alpha_k + G_B\beta_k \ , \qquad (k = 1, \ldots, K) \ . \qquad (13\text{–}17)$$

Let us suppose y_k $(k = 1, \ldots, K)$ is perfectly fit by the SAM but not by the WAM. (If it can be perfectly fit by the WAM, it can be perfectly fit by the SAM separately, but not vice versa.) The relationship between $G_A\alpha_k$ and $G_A\alpha w_{kA}$ is depicted in Figure 13–6. The vector, $G_A\alpha w_{kA}$, represents the portion of $G_A\alpha_k$ that can be explained by the WAM. In other words, $G_A\alpha_k$ is decom-

Figure 13–6. The Relationship between the WAM and the SAM



posed into two parts: $G_A\alpha w_{kA}$, the part that can be accounted for by the WAM; and $G_A\alpha_k - G_A\alpha w_{kA}$, the part that cannot be accounted for by the WAM. Furthermore, $G_A\alpha w_{kA}$ and $G_A\alpha_k - G_A\alpha w_{kA}$ are mutually orthogonal so that:

$$SS(G_A\alpha_k) = SS(G_A\alpha w_{kA}) + SS(G_A\alpha_k - G_A\alpha w_{kA}) , \qquad (13\text{–}18)$$

where SS(**v**) is the sum of squares of elements of vector **v**. The same relation holds for all k' and also for other factors in the models.

When there are only two levels in each factor, the WAM and the SAM are compatible in every way. In this case, α_k can always be written as αw_{kA}.

Relations to Other Models

We have explained the relationship between the WAM and the SAM. We have also mentioned that when the independent variables are completely specified—that is, when they are a priori quantified—the WAM reduces to regression analysis (linear or monotonic, depending on the scale level of the dependent variables). Examples of this class of models are Carroll's (1972) external unfolding analysis and Srinivasan and Shocker's (1973) composite criterion model.

The idea of describing differences between sets of observations by a differential weighting of additive effects is apparently not new. In fact, most of the multivariate data analysis methods can be thought of as techniques to find various optimal linear combinations of variables according to some prescribed criteria (McDonald 1968). In particular, when the independent variables (factors) are completely unspecified in the WAM, we obtain the bilinear model

$$y_{kr} = \sum_{t=1}^{T} w_{kt} f_{rt} \ , \qquad (13\text{–}19)$$

where y_{kr} is the model prediction for the kth observation set (usually subject k) on variable r, f_{rt} is the coordinate (conventionally called factor loading) of variable r on factor t, and w_{ka} is the weight (or factor score) of subject k on factor a. Since f_{rt} and w_{kt} are assumed to be unknown, in this case, both w_{kt} and f_{rt} should be estimated under some identifiability constraints. The number of factors, T, is the minimum number of hypothetical attributes (latent traits, factors) necessary to approximate the observed data to a sufficient degree and is also to be determined in the course of data analysis. Under a specialized set of identifiability constraints, the bilinear model is called the *principal component analysis*. In a sense, f_{rt} is a weight attached to w_{kt} (responsible for the variation of y_{kr} over r), just as w_{kt} is a weight attached to f_{rt} (responsible for the variation of y_{kr} over k). The roles of w_{kt} and f_{rt} are dual in this sense.

A special case of (13–19) has been proposed by Carroll and Chang (1970) in order to describe subject differences in dissimilarity judgments. The proposed model—called INDSCAL—is the weighted distance model, which is written as:

$$d_{kij}^2 = \sum_{t=1}^{T} w_{kt}(x_{it} - x_{jt})^2 \ , \qquad (13\text{–}20)$$

where d_{kij} is the Euclidean distance between stimuli i and j for subject k, x_{it} is the coordinate of stimulus i on dimension t, and T is the dimensionality of the space in which a set of stimuli are represented. That (13–20) is a special case of (13–19) can be readily seen by noting that

$$f_{rt} = (x_{it} - x_{jt})^2 \ , \qquad (13\text{–}21)$$

where stimulus pair (i,j) is indexed by r. That is, INDSCAL is a special type of the WAM in which squares of the dimensionwise differences between two coordinates are taken as additive effects. The fact that (13–20) is a special case of the WAM provides an interesting application of the WAM to similarity and dissimilarity data. In (13–20), the weights are usually restricted to be nonnegative. This is because both $(x_{it} - x_{jt})^2$, $(t = 1, \ldots, T)$ and d_{kij}^2 are nonnegative. A minor generalization of (13–20) is

$$d^{p}_{kij} = \sum_{t=1}^{T} w_{kt} |x_{it} - x_{jt}|^{p} , \qquad (13\text{–}22)$$

where p is the Minkowski power ($1 \leq p$).

When f_{rt} in (13–19) is constrained to be

$$f_{rt} = \sum_{q_t} h_{rq_t} \alpha_{q_t} ,$$

we obtain (13–3), which is the general form of the WAM.

PARAMETER ESTIMATION

In this section, we describe methods for parameter estimation in the WAM. In the previous section, the basic properties of the WAM were explored strictly within the context of the model. However, data analysis is an interplay between the model and the data. Thus, we first need to discuss various data characteristics to be taken into account in developing parameter estimation procedures.

The Data

The necessity of discussing data characteristics stems from the fact that the model predictions are usually not directly observable. The observed data typically contain a sizable amount of errors. In addition, some systematic distortions may be under effect in a measurement process. For example, if the data are collected by a rating scale, they may be distorted monotonically (relative to the model predictions) towards the ends of the scale due to possible floor and ceiling effects. The observed data in this case will not possess interval-scale properties. In this small section, we briefly discuss some of the data characteristics to be considered in fitting the model. These considerations are not restricted to the WAM, and more comprehensive treatments of this topic can be found in Young (1981) and Takane (1982).

We first discuss the distinction made in terms of scale levels of measurement. (This is basically the distinction made by WADDALS, a least-squares procedure for fitting the WAM.) The simplest case is the one in which the data are obtained on an interval scale or higher. In this case, it is assumed that y_{kij} is only error-perturbed and that no systematic distortion is exerted on the error-perturbed model prediction. That is,

$$y^{*}_{kij} = y_{kij} + e_{kij} , \qquad (13\text{–}23)$$

where e_{kij} is an error random variable and y^{*}_{kij} is directly observable. The model fitting in this case only involves the estimation of parameters in the WAM (namely, the additive effects and the differential weights). Most often, the estimation is accomplished by minimizing the overall size of e_{kij} (that is,

$\Sigma\, e_{kij}^2$) without further assuming any distributional properties on e_{kij}.

The interval information contained in y_{kij}^* may get lost to varying degrees in an observation process. Some monotonic distortion may be in effect. For example, the subject may report only rank orders among y_{kij}^*. Thus, only ordinal information is preserved. The data are said to be measured on an ordinal scale in this case. The ordinal data have to be monotonically transformed back to y_{kij}^* before they can be subjected to a representation by the model. Specifically,

$$f_m(o_{kij}) = y_{kij}^* = y_{kij} + e_{kij} \quad , \tag{13–24}$$

where o_{kij} is the observed data and f_m is some monotonic function. The model fitting in this case involves a monotonic transformation of the observed data as well as the model estimation. Again this is most often done by minimizing the overall size of e_{kij}. The WADDALS algorithm to be described in the next section is primarily aimed at ordinal data, although it can be used for other data as well. The ordinal scale assumption is sufficiently flexible, so that almost any data can be analyzed under this assumption (with an exception to be described shortly). If the data happen to be measured at a higher scale level, an approximately linear transformation is usually found for f_m, which is indeed the correct transformation in this case. One should note, however, that some efficiency is bound to get lost if one makes a weaker measurement assumption than needs to be made.

In certain extreme cases, only nominal information may be obtained. At the nominal scale level, two observations are either indifferent or distinct. For example, we may ask the subject to classify y_{kij}^* into several unordered categories. Observations that are in the same category are considered indifferent, while those in different categories are considered distinct. Nominal observation categories have to be quantified before the data can be represented by the model. This quantification is sometimes called a *nominal transformation* and can be done just as the levels of the additive factors are quantified in the WAM.

In all the above cases, the data can be arranged in a matrix form for each set (assuming that there are only two factors) in the same way as the model predictions are arranged in Figure 13–2. In the interval case, y_{kij} is replaced by y_{kij}^*, and in the ordinal case, it is replaced by o_{kij} (which is assumed monotonically related to y_{kij}^*). In the nominal case, it is replaced by the category number into which y_{kij}^* is classified. Multiple sets of observations, each arranged in a matrix form, compose a multimode data set.

The data characteristics discussed above are based on the scale levels of measurement. However, in certain situations, other characteristics of the data may be more meaningful. For example, there are many different ways of collecting the data, and each method may require a different mental process by which the observed data are generated. If a model for this process can be constructed, the observed data can be more directly related to

its unobservable counterpart, the model. Takane (1982) incorporated a specific process model for each of three representative types of data collection methods—categorical ratings, pair comparisons, and directional rank orders—in fitting the additive model. We briefly discuss each of these data collection methods in turn.

In the categorical rating method, the subject is asked to rate y^*_{kij} on a rating scale having a relative small number of observation categories (say, up to 7 or 9). Thus, only category membership of y^*_{kij} is observed. Since categories are usually ordered, the data may be analyzed as mere ordinal data. Alternatively, a specific process model of categorical judgments may be incorporated. Through this model, and under some distributional assumption on e_{kij}, the probability of a certain categorical observation can be derived as a function of model parameters. The categorical rating data may be represented in a matrix form with each element in a matrix representing a category number. When many replicated observations are made for each y^*_{kij}, it may be economical to represent the data in a frequency form, in which observed frequencies of categorical observations are given along with stimulus indices, i and j, and k.

In the pair comparison method, the subject is presented two stimuli at a time and asked to judge which one "dominates" the other according to some prescribed criterion. Thus, only pairwise ordinal information about y^*_{kij} and $y^*_{ki'j'}$ (that is, either $y^*_{kij} > y^*_{ki'j'}$ or $y^*_{kij} < y^*_{ki'j'}$) is obtained. The pair comparison data may be analyzed as partial rank-order data. On the other hand, a specific process model for pair comparison judgments may be constructed. The probability of $y^*_{kij} > y^*_{ki'j'}$ or $y^*_{kij} < y^*_{ki'j'}$ can then be derived as a function of model parameters. The pair comparison data cannot be represented in a matrix form, since two stimuli are involved in each judgment. This type of data can be represented most naturally in a frequency form in which frequencies of comparative judgments are provided along with stimulus indices, i and j, and k.

In the directional rank-order method, the subject is asked to rank order a set of stimuli in a specified direction—either from the largest to the smallest, or vice versa. All stimuli generated by factorial combinations of the levels in all factors may be rank ordered in this way. It is important that the ranking is made in a specified direction. Otherwise, it is extremely difficult to construct a process model that relates y^*_{kij} to an observed ranking. Under the directionality of the ranking process, the probability of observed rank orders can be derived in a relatively straightforward manner. The directional rank-order data may be represented in a matrix form with rank numbers as entries of matrices. The frequency representation is usually not convenient, since almost all rankings are unique, unless the number of stimuli to be rank ordered is relatively small. The directional rank-order data can also be analyzed as mere ordinal data.

A category response model may also be constructed for nominal data. For example, the log-linear model (Andersen 1980; Bishop, Fienberg, and Holland 1975; Bock 1975; Goodman 1978) for analysis of contingency tables can be thought of as the SAM with the logistic response function (Luce's [1959] choice model) and the

multinomial probability model for unordered categorical responses. It may be interesting and worthwhile to extend the log-linear model using the WAM. (See the application section for more detail.)

A Brief Overview of the Existing Methods for Additivity Analysis

A number of estimation procedures are available for the SAM. When the data are measured on an interval scale or higher, the SAM reduces to an ANOVA model without interaction terms, which is in turn a special case of the general linear regression model with dummy independent variables (Draper and Smith 1981). A number of procedures have been developed for ordinal data—MONANOVA (Kruskal 1965), ADDIT (Roskam 1968), ADDALS (de Leeuw, Young, and Takane 1976), and MORALS (Young, de Leeuw, and Takane 1976)—which all use Kruskal's (1964a, 1964b) least-squares monotonic transformation as part of their algorithm. CM-I (Conjoint Measurement I) by Lingoes (1973), on the other hand, uses Guttman's rank-image transformation (Guttman 1968). A procedure by Ramsay (1977) and the one by Winsberg and Ramsay (1980) use smoother transformations such as power and spline transformations. The nominal data lead to a special case of discriminant analysis in which independent variables are also categorical. Hayashi's (1952) Quantification Method II and Carroll's (1973) CCM (Categorical Conjoint Measurement) have been developed specifically for this situation. However, by appropriate codings of dummy variables, the same analysis could be performed by canonical correlation analysis.

Fitting procedures are much more limited for the WAM. To the best of my knowledge, WADDALS and MAXADD (Takane et al. 1980; Takane 1982) are the only two that are available at the moment. MAXLIN is now under development, which generalizes MAXADD to the general weighted linear (not restricted to additive) model. De Sarbo, Carroll, Lehmann, and O'Shaughnessy (forthcoming) recently proposed a three-way multivariate conjoint analysis method, which, as its special case, subsumes the WAM. Their procedure, however, is restricted to interval-scale data.

In this section, we discuss WADDALS and MAXADD in some detail. An excellent review of the algorithms for the SAM has been provided by Rao (1977).

ADDALS

As an introduction to WADDALS, we first discuss the ADDALS algorithm (de Leeuw et al. 1976; Young 1981) for the SAM. WADDALS works within the same algorithmic framework as ADDALS, known as alternating least-squares.

Let us write the SAM in a vector form for a single set of observations, analogous to (13–16) and (13–17). Let $\mathbf{y}$ denote the vector of model predictions:

$$\mathbf{y} = G_A\alpha + G_B\beta \ . \qquad (13\text{–}25)$$

This vector is fitted to the vector of error-perturbed model predictions, which is denoted by $\mathbf{y}^*$. We primarily focus on the case in which $\mathbf{y}^*$ is measured on an ordinal scale. (Later, we will discuss whatever modifications are necessary for the nominal case.) We would like to estimate both model parameters—α and β—and a monotonic function—f_m in (13–24)—in such a way that the discrepancy between $\mathbf{y}^*$ and y is a minimum. This may be expressed as:

$$\phi = (\mathbf{y}^* - \mathbf{y})' (\mathbf{y}^* - \mathbf{y}) / \mathbf{y}^{*\prime}\mathbf{y}^* , \tag{13–26}$$

which is to be minimized with respect to α, β, and $\mathbf{y}^*$, where $\mathbf{y}^*$ is monotonically related to $\mathbf{o}$. ADDALS minimizes this by alternating the following two least-squares steps, until successive values of ϕ exhibit no significant improvement. The positive square root of ϕ is called *stress*.

Step 1. Model Estimation:
Obtain $\mathbf{y}$ *that minimizes* ϕ *for fixed* $\mathbf{y}^*$ *(such that* $\mathbf{y}^{*\prime}\mathbf{y}^* = 1$*).*

Step 2. Optimal Scaling:
Obtain $\mathbf{y}^*$ *that minimizes the numerator of* ϕ *for fixed* $\mathbf{y}$*, and normalize* $\mathbf{y}^*$ *so that* $\mathbf{y}^{*\prime}\mathbf{y}^* = 1$.

In the initial iteration, $\mathbf{o}$ is centered and normalized and used for $\mathbf{y}^*$. In Step 1, $\mathbf{y}$ is obtained by obtaining the least-squares estimates of α and β, which are:

$$\hat{\alpha} = (G'_A G_A)^{-1} G'_A \mathbf{y}^* \tag{13–27}$$

and

$$\hat{\beta} = (G'_B G_B)^{-1} G'_B \mathbf{y}^* .$$

If the elements $\mathbf{y}^*$ are rearranged in a matrix form so that its (i,j)th element is y^*_{ij} corresponding to y_{ij} in (13–13), $\hat{\alpha}$ and $\hat{\beta}$ are just the vectors of row and column means, respectively, of this matrix. The least-squares estimate $\mathbf{y}^*$ in Step 2 is obtained by applying Kruskal's least-squares monotonic regression algorithm (Kruskal 1964b) to $\mathbf{y}$.

The above iteration scheme has been proven to be convergent (de Leeuw et al. 1976). This may be easily seen by noting that ϕ can never get larger, since conditional least-squares estimates are obtained in each step. Furthermore, ϕ is bounded below ($\phi \geq 0$). By a standard theorem, in functional analysis, a monotonically decreasing sequence is bound to converge. The convergence point may not be a local minimum of ϕ in general, but this is guaranteed by the continuity of ϕ (Kruskal 1971).

That Step 2 really minimizes ϕ with respect to $\mathbf{y}^*$ is assured by the following theorem, which is stated without proof:

THEOREM 13–1. *On the minimization of normalized least-squares criteria.*

Let $Q(\mathbf{a}) = \mathbf{a}'\mathbf{a}$,

$$\phi^*(\mathbf{x}_1,\mathbf{x}_2) = Q[\mathbf{y}(\mathbf{x}_1) - \mathbf{z}(\mathbf{x}_2)],$$

$$\phi_y(\mathbf{x}_1,\mathbf{x}_2) = \phi^*(\mathbf{x}_1,\mathbf{x}_2)/Q[\mathbf{y}(\mathbf{x}_1)],$$

and

$$\phi_z(\mathbf{x}_1,\mathbf{x}_2) = \phi^*(\mathbf{x}_1,\mathbf{x}_2)/Q[\mathbf{z}(\mathbf{x}_2)].$$

1. *Let* $\mathbf{x}_1^{(1)}$ *and* $\mathbf{x}_2^{(1)}$ *minimize* ϕ_y.
2. *Let* $\mathbf{x}_1^{(2)}$ *and* $\mathbf{x}_2^{(2)}$ *minimize* ϕ^* *subject to the restriction that* $Q[\mathbf{y}(\mathbf{x}_1^{(2)})] = 1$.
3. *Let* $\mathbf{x}_1^{(3)}$ *and* $\mathbf{x}_2^{(3)}$ *minimize* ϕ^* *subject to the restriction that* $Q[\mathbf{z}(\mathbf{x}_2^{(3)})] = 1$.
4. *Let* $\mathbf{x}_1^{(4)}$ *and* $\mathbf{x}_2^{(4)}$ *minimize* ϕ_z.

The four sets of solutions are essentially equivalent. That is, $\mathbf{x}_1^{(1)} \sim \mathbf{x}_1^{(4)}$ *are proportional to each other, and so are* $\mathbf{x}_2^{(1)} \sim \mathbf{x}_2^{(4)}$.

The denominator of ϕ in (13–26) is just a normalization factor for preventing ϕ from converging to 0 by shrinking the size of $\mathbf{y}$ and $\mathbf{y}^*$. The above theorem shows that it does not matter by which, $\mathbf{y}$ or $\mathbf{y}^*$, the normalization is done and that the normalization is effected by first obtaining $\mathbf{y}^*$, which minimizes the numerator of ϕ, and then by actually normalizing this $\mathbf{y}^*$ so that $\mathbf{y}^{*\prime}\mathbf{y}^* = 1$.

The above algorithm is not completely free from difficulties. There are problems of convergence to a nonglobal minimum, nonuniqueness of the minimum reached, and a possible degeneracy of solutions. These problems are not restricted to the ADDALS algorithm but are discussed here only because it happens to be the first algorithm discussed in this chapter.

The convergence point attained by ADDALS may not be the global minimum of ϕ. Only a convergence to a local minimum is ensured. There is no definitive way to get over this problem. Good initial starts often help, but there is no guarantee that they always work. It is thus advisable to obtain several solutions from different initial starts, and if they do not all agree, choose the one that gives the smallest value of ϕ.

No general conditions have been established for the uniqueness of the parameter estimates. The local minimum attained may not be a strong minimum but merely a weak local minimum in the sense that we may find $\hat{\mathbf{y}}^*$ and $\hat{\mathbf{y}}$, such that $\phi(\mathbf{y}^*, \mathbf{y}) = \phi(\hat{\mathbf{y}}^*, \hat{\mathbf{y}})$ sufficiently near $\mathbf{y}^*$ and $\mathbf{y}$, but $\hat{\mathbf{y}}^* \neq \mathbf{y}^*$ and $\hat{\mathbf{y}} \neq \mathbf{y}$. A general recommendation is to obtain many observations. With an increasing number of observations (relative to the number of parameters to be estimated), the seriousness of the problem decreases. As more observations are made, the region of $\mathbf{y}^*$ and $\mathbf{y}$, in which ϕ

is equal, gets smaller in size, until it can be virtually considered as a unique point. The best way to obtain more observations, without increasing the number of model parameters, is to obtain replicated observations.

A solution is said to be degenerate when $\phi = 0$ and the elements of $\mathbf{y}^*$ split into two or more equivalence groups. A degenerate solution occurs when rows or columns are not ordinally connected in the following sense:

DEFINITION 13–1. *Order Connectedness*

Rows (columns) of a data matrix are said to be ordinally connected when in each row (column) there is at least one element that is larger than at least one element in the other rows (columns) and there is at least one element that is smaller than at least one element in the other rows (columns).

Rows and columns cannot be unconnected simultaneously. The following simple numerical example (3 × 3) shows what happens when rows or columns are unconnected:

9	8	7
6	5	3
4	2	1

Rows of this data split into two groups between the first and second rows (indicated by the line). Rows are unconnected here. These data can be perfectly fit by:

	α_j		
	0	0	0
2	2	2	2
β_j			
-1	-1	-1	-1
-1	-1	-1	-1

The model predictions split into two groups in the same way as the data split. Since monotonicity is perfectly satisfied in this case, we obtain $\phi = 0$. Although the fit is perfect, the solution is trivial.

Note that the above data satisfy the independence condition, one of the necessary conditions for an additive representation. In fact, "too clean" data (the data that can be perfectly represented by the additive model) are less likely to be ordinally connected than moderately noisy data and are thus not particularly suitable for ordinal additivity analysis. For ADDALS to work it is necessary that the data contain a moderate degree of error. The best way to avoid the degenerate solution is to

increase the number of levels so that the levels are sufficiently close to generate "confusions" among rows and columns. Replicated observations are also helpful in creating ordinal connectedness in the data.

Some modifications are necessary to the standard ADDALS algorithm when the design is not balanced and/or the data are not ordinal. When there are missing observations, we may make the design artificially balanced by assigning them an arbitrary value (say, zero) initially and by assigning model predictions subsequently. This treatment is not described in the original ADDALS paper (de Leeuw et al. 1976), but this approach has been taken in the WADDALS algorithm.

When the data are nominal, the quantification of the observation categories may be expressed as:

$$\mathbf{y}^* = \mathbf{S}\boldsymbol{\Theta} \ , \tag{13–28}$$

where S is a matrix of dummy variables. Specifically, $\mathbf{S} = [s_{ij}]$ where

$$s_{ij} = \begin{cases} 1, \text{ if stimulus } i \text{ is responded by category } j, \\ 0, \text{ otherwise,} \end{cases}$$

and $\boldsymbol{\Theta}$ is the quantification vector of the observation categories. The least-squares estimate of $\boldsymbol{\Theta}$ is obtained by:

$$\hat{\boldsymbol{\Theta}} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{y} \ . \tag{13–29}$$

Kruskal's monotonic regression algorithm in Step 2 should be replaced by (13–28) with $\hat{\boldsymbol{\Theta}}$ obtained by (13–29). The iterative procedure in this case is known to be equivalent to obtaining eigenvectors of a certain matrix, which is exactly what is done by Quantification Method II or CCM.

WADDALS

Having discussed the ADDALS algorithm in the previous section, there is only one more step necessary to arrive at the WADDALS algorithm (Takane et al. 1980), in which an additional set of parameters, namely the differential weights, have to be estimated. WADDALS minimizes the least-squares criterion:

$$\phi_w = \sum [(\mathbf{y}_k^* - \mathbf{y}_k)'(\mathbf{y}_k^* - \mathbf{y}_k)/\mathbf{y}_k^{*\prime}\mathbf{y}_k^*] \ . \tag{13–30}$$

Note that ϕ_w is essentially the sum of individual least-squares criteria, like the one for ADDALS ([13–26]), summed over k. The positive square root of mean ϕ_w ($\sqrt{\phi_w/K}$) is called *stress* in WADDALS. The WADDALS algorithm is again based on the alternating least-squares principle. It minimizes ϕ_w alternately with respect to model parameters (α, β, w_{kA}, w_{kB}) and with respect to

data transformations ($\mathbf{y}_k^*$). The model estimation phase now consists of two subphases, one estimating the weights and the other estimating the additive effects. So the basic iterative cycle is:

Step 1. Model Estimation:

Obtain $\mathbf{y}_k$ ($k = 1, \ldots, K$) that minimizes ϕ_w for fixed $\mathbf{y}_k^*$ (such that $\mathbf{y}_k^{*\prime}\mathbf{y}_k^* = 1$ for $k = 1, \ldots, K$). This consists of the following two substeps.

1. *Estimate the weights (w_{kA}, w_{kB}) for $k = 1, \ldots, K$ for fixed $\mathbf{y}_k^*$ and for fixed additive effects.*
2. *Estimate the additive effects (α, β) for fixed $\mathbf{y}_k^*$ and for fixed weights.*

Step 2. Optimal Scaling:

Obtain $\mathbf{y}_k^*$ ($k = 1, \ldots, N$) that minimizes ϕ_w for fixed $\mathbf{y}_k$ and normalize them so that $\mathbf{y}_k^{*\prime}\mathbf{y}_k^* = 1$ for $k = 1, \ldots, K$.

In the initial iteration, we set $\mathbf{y}_k^* = \mathbf{o}_k$ ($k = 1, \ldots, K$), which are to be centered and normalized. Initial estimates of the additive effects are obtained by applying the SAM, treating k as a replication. They are then normalized so as to satisfy (13–11).

The weights are estimated as follows. Let

$$W = \begin{bmatrix} w_{1A} & w_{1B} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ w_{KA} & w_{KB} \end{bmatrix},$$

$$\mathbf{X} = [\mathbf{G}_A\hat{\alpha}, \mathbf{G}_B\hat{\beta}],$$

and

$$\mathbf{Y}^* = [\mathbf{y}_1^*, \ldots, \mathbf{y}_K^*].$$

Then,

$$\hat{\mathbf{W}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}^*.$$

The least-squares estimates of the additive effects are obtained by

$$\hat{\alpha} = (G'_{AW} G_{AW})^{-1} G'_{AW} Y^* ,$$

for α, where $Y^{*\prime} = (Y_1^{*\prime}, \ldots, Y_K^{*\prime})$, and

$$G_{AW} = \begin{bmatrix} w_{1A} & G_A \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ w_{KA} & G_A \end{bmatrix} .$$

The estimate of β is similarly obtained. Once $\hat{\alpha}$ and $\hat{\beta}$ are obtained, they are normalized to satisfy (13–11), and the weights are adjusted for the normalization. Then the model predictions are calculated:

$$Y = [y_1, \ldots, y_N] = XW .$$

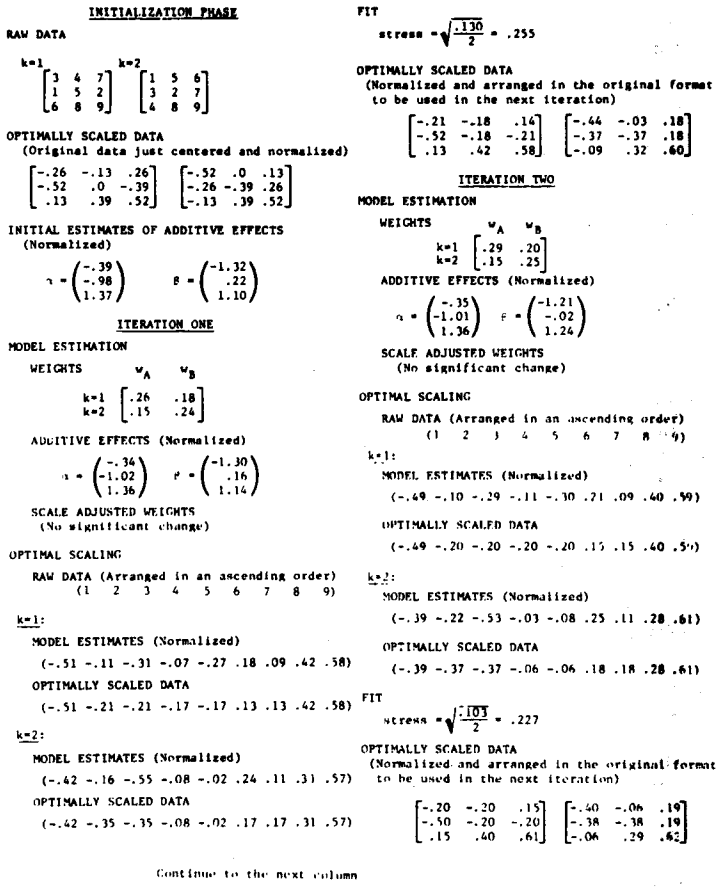
Finally, y_k^* ($k = 1, \ldots, K$) are obtained by applying Kruskal's monotonic regression to each y_k separately. (If the data are nominal, the least-squares nominal transformation is applied instead.)

The two phases are repeatedly applied until convergence is attained. The convergence is assumed to have been reached when the reduction in ϕ_w from one iteration to the next becomes negligible (for instance, 0.0005). The first two iterations of the WADDALS algorithm are shown in Figure 13–7 for hypothetical sets of data. The reader is encouraged to follow through the iterations in order to obtain a concrete image of how WADDALS works. Similar example iterations are shown for ADDALS in Young (1981).

In Figures 13–8 and 13–9, the plots of estimated α_{kj} and β_{kj} from ADDALS and those of estimated $w_{kA}\alpha_j$ and $w_{kB}\beta_j$ from WADDALS are shown. The data analyzed were the hypothetical data sets given in Figure 13–7. Figure 13–8 indicates that the data are not perfectly suitable for the WAM. However, the design is very small (3×3); the estimates are based on only three observations each. Since these estimates are not likely to be very reliable, it may still be worthwhile to apply the WAM. Figure 13–9 exhibits the characteristic patterns of proportionality among $w_{kA}\alpha_j$ and $w_{kB}\beta_j$. (Compare them with Figure 13–1.)

Stress values obtained from separate ADDALS analyses of the two sets are .198 for $k = 1$ and .194 for $k = 2$. The joint stress is computed by $[\{(.198)^2 + (.194)^2\}/2]^{1/2}$, which turns out to be .196. The stress value obtained from WADDALS is .208, a slight

Figure 13–7. Example of WADDALS Iterations



Continue to the next column

increase from .196. The difference represents the portion of variability in the data that can be represented by the SAM but not by the WAM. Whether the difference is statistically significant or not is difficult to determine; there is no definite criterion, but when the design is as small as this, a difference of .012 seems quite large.

The comments made for ADDALS concerning the nonglobal minimum, the weak minimum, and the degeneracy of solutions are also valid for WADDALS. The last two of these problems are less serious in WADDALS, since the number of observations to number of parameters ratio is generally much larger in the WAM.

A necessary condition to avoid a degenerate solution is much weaker for the WAM. Ordinally connected rows (or columns) define a partition of rows (columns) in the data matrix. Within each subset in the partition, rows (columns) are connected, while rows (columns) in different subsets are not connected. When the

Figure 13–8. Estimates of Additive Effects in SAM Obtained from Hypothetical Data Sets

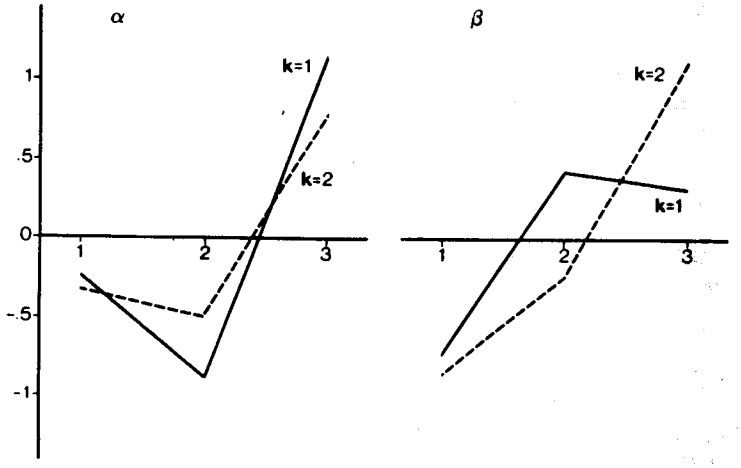
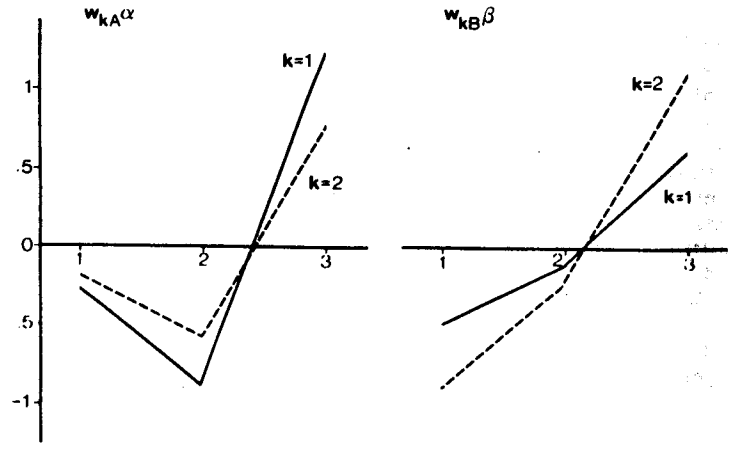


Figure 13–9. Estimates of Additive Effects Multiplied by Weights in the WAM*



*Derived from the same hypothetical data sets as in Figure 13–8.

partition consists of only one set of all rows (columns), the data are said to be rowwise (columnwise) connected.

DEFINITION 13–2

Two subsets of connected rows (columns) are connected if at least one element in one subset is larger and at least one

element in the same set is smaller than at least one element in the other subset.

Unconnected subsets for some k should be connected for at least one $k'(\neq k)$.

MAXADD

WADDALS, as discussed in the previous section, works reasonably well as a descriptive device. It is better suited for obtaining crude estimates based on relatively weak assumptions (monotonicity, no distributional assumptions). However, it can provide very little information pertinent to model evaluations, such as assessing the reliability of the estimates, comparing the goodness-of-fit between two models, and so forth. The MAXADD procedure developed by Takane (1982), on the other hand, allows various statistical inferences that have not been possible previously.

The MAXADD procedure is capable of fitting both the SAM and the WAM to categorical rating, pair comparison, and directional rank-order data. In MAXADD, nonmetric data are viewed as incomplete data (Dempster, Laird, and Rubin 1977). A metric process conveying complete information about the model is assumed to underlie such nonmetric data. The metric information is presumed to get lost in an observation process, leaving us only ordinal or nominal information. A specific information reduction mechanism is postulated for each specific type of data; through this mechanism, the likelihood of observed nonmetric data is stated as a function of model parameters. Model parameters are then determined so as to maximize the likelihood of the observed data. Procedures similar to MAXADD have been developed for multidimensional scaling (Takane 1978, 1981; Takane and Carroll 1981).

We discuss the model linking the unobserved metric process to the observed nonmetric data for each of the three data collection methods mentioned above. In the categorical rating method, each category is represented by an interval specified by its upper and lower boundaries. Let the upper and lower boundaries of the mth observation category be denoted by b_{km} and $b_{k(m-1)}$, respectively. We assume that y_{kij} is put into category m whenever its corresponding error-perturbed model prediction falls in an interval bounded by $b_{k(m-1)}$ and b_{km}. Let p_{kijm} denote the probability of this event. Then

$$p_{kijm} = \text{Prob}\ (b_{k(m-1)} < y^*_{kij} < b_{km}) \ . \tag{13-31}$$

Define

$$F_{kijm} = \text{Prob}\ (y^*_{kij} < b_{km}) \ . \tag{13-32}$$

Then $p_{kijm} = F_{kijm} - F_{kij(m-1)}$. Under the distributional assumption that y^*_{kij} is logistic, (13–32) can be more explicitly written as:

$$F_{kijm} = [1 + \exp\{-s_k\ (b_{km} - y_{kij})\}]^{-1} \ , \tag{13-33}$$

where s_k is a dispersion parameter. Let Z_{kijm} denote the observed frequency with which y_{kij} is put into category m. The joint probability (p_{kij}) of Z_{kijm} ($m = 1, \ldots, M$) can be stated as:

$$p_{kij} = \prod_{m=1}^{M} (p_{kijm})^{Z_{kijm}} . \qquad (13\text{–}34)$$

Finally, the joint probability of the total set of category judgments is obtained by the product of the above p_{kij} over i, j, and k, assuming that the judgments are all statistically independent.

In the pair comparison case, we may assume that y_{kij} is judged to be larger than y_{kuv} whenever y^*_{kij} exceeds y^*_{kuv}. The probability (p^*_{kijuv}) that y_{kij} is judged larger than y_{kuv} is then written as:

$$\begin{aligned} p^*_{kijuv} &= \text{Prob}\ (y^*_{kij} > y^*_{kuv}) \qquad (13\text{–}35) \\ &= \text{Prob}\ (y^*_{kij} - y_{kuv} > 0) , \end{aligned}$$

which, assuming that $y^*_{kij} - y^*_{kuv}$ is logistic, can be more explicitly stated as:

$$p^*_{kijuv} = [1 + \exp\{-s_k(y_{kij} - y_{kuv})\}]^{-1} . \qquad (13\text{–}36)$$

Let Z_{kijuv} denote the frequency with which y_{kij} is judged larger than y_{kuv}. Then the joint probability (p_{kijuv}) of Z_{kijuv} and Z_{kuvij} can be stated as:

$$p_{kijuv} = (p^*_{kijuv})^{Z_{kijuv}} (1 - p^*_{kijuv})^{N_{kijuv} - Z_{kijuv}} , \qquad (13\text{–}37)$$

where $N_{kijuv} = Z_{kijuv} + Z_{kuvij}$. The joint probability of the total set of pair comparison judgments is obtained by the product of p_{kijuv}.

In the directional rank-order method, each ranking is performed in a specified direction, either from the largest to the smallest or from the smallest to the largest. In this case, we may assume that each ranking is obtained by successive first choices. Suppose that the ranking is performed from the largest to the smallest among M stimuli. Let $y_k^{(m)}$ denote the model prediction judged to be the mth largest, and $y_k^{(m)*}$ the corresponding error-perturbed model prediction. We assume that when $y_k^{(m)}$ is chosen as the mth largest element, the $m - 1$ successive first choices have already been made and the corresponding $m - 1$ stimuli have been deleted from the comparison set. We assume that $y_k^{(m)}$ is chosen as the largest element from the remaining $M - m + 1$ stimuli. We also assume that this event occurs whenever $y_k^{(m)*}$ exceeds all other error-perturbed model predictions remaining in the comparison set. Then

$$p_k^{(m)} = \text{Prob}\ (y_k^{(m)*} > y_k^{(m+1)*}, \ldots, y_k^{(m)*} > y_k^{(M)*}) \quad , \quad (13\text{–}38)$$

where $p_k^{(m)}$ is the probability that $y_k^{(m)}$ is chosen as the largest element among the $M - m + 1$ elements. Using Luce's (1959) choice model, $p_k^{(m)}$ can be more explicitly stated as:

$$p_k^{(m)} = \left[1 + \sum_{j=m+1}^{M} \exp\ \{-s_k (y_k^{(m)} - y_k^{(j)})\}\right]^{-1} \quad . \quad (13\text{–}39)$$

Assuming that each successive choice is performed independently, the probability of a ranking is obtained by the product of $p_k^{(m)}$ over $m = 1, \ldots, M - 1$. The joint probability of multiple rankings, on the other hand, can be stated as the product of the probabilities of individual rankings. A rationale and the condition for the statistical independence of successive first choices are given in Takane and Carroll (1981).

In all cases, the likelihood function is stated as a function of model parameters. Their estimates are determined in such a way as to maximize the likelihood function. Let L denote the likelihood function. Then, we would like to solve likelihood equations,

$$\frac{\partial \ln L}{\partial \boldsymbol{\theta}} = \mathbf{0} \ ,$$

for $\boldsymbol{\theta}$, the parameter vector. MAXADD uses the Fisher scoring algorithm for solving the likelihood equations. This algorithm, starting from some initial estimates, updates the parameter estimates by solving

$$\varepsilon^{(q)}\, \mathbf{I}(\boldsymbol{\theta}^{(q)})(\boldsymbol{\theta}^{(q+1)} - \boldsymbol{\theta}^{(q)}) = \mathbf{u}(\boldsymbol{\theta}^{(q)}) \quad (13\text{–}40)$$

for $\boldsymbol{\theta}^{(q+1)}$, where $\varepsilon^{(q)}$ is a step-size parameter, $\boldsymbol{\theta}^{(q+1)}$ and $\boldsymbol{\theta}^{(q)}$ are new and old parameter estimates, respectively, and

$$\mathbf{u}(\boldsymbol{\theta}) = \left(\frac{\partial \ln L}{\partial \boldsymbol{\theta}}\right)$$

and

$$\mathbf{I}(\boldsymbol{\theta}) = E\left[\left(\frac{\partial \ln L}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln L}{\partial \boldsymbol{\theta}}\right)'\right] \quad ,$$

which is called *Fisher's information matrix*. Updating of equation (13–40) is iteratively applied until convergence is reached. Initial estimates are obtained by the same procedure used in ADDALS and WADDALS, except for pair comparison data. In the pair comparison case, estimates of y_{kij} are first obtained by applying a least-squares procedure for the ordinary Luce model to observed frequencies, Z_{kijuv}, and then these estimates are used to obtain initial estimates of model parameters. The reader is referred to Takane (1982) for more details of the algorithm.

As has been emphasized, one of the major advantages of the MAXADD procedure is its statistical inference capability. The Moore-Penrose inverse of the information matrix evaluated at the maximum likelihood estimates of parameters is known to give asymptotic variance-covariance estimates of the parameter estimates (Ramsay 1978). These estimates provide information concerning the reliability of the derived estimates.

The asymptotic chi-square statistic derived from the likelihood ratio principle may be used for various model comparisons. Let L_0 and L_1 represent the two likelihoods obtained under hypothesis H_0 (which postulates a more restricted model) and hypothesis H_1 (which postulates a less restricted model), respectively. Then

$$\chi^2 = 2(\ln L_1 - \ln L_0)$$

follows asymptotically the chi-square distribution with degrees of freedom equal to the difference in the number of parameters in the two models. Observed values of χ^2 are compared with appropriate critical values of chi-square to test a significant difference between the two models.

The asymptotic chi-square can only be used when one of the two models compared is a special case of the other. When this is not the case, the following statistic will be helpful:

$$AIC(\pi) = -2 \ln (L_\pi) + 2n_\pi ,$$

(Akaike 1974), where L_π is the maximum likelihood of model π and n_π is the effective number of parameters in the model. A smaller value of AIC indicates a better fit.

Some caution should be exerted when one wants to rely on the statistical inference features of MAXADD. These are all based on the asymptotic properties of the maximum likelihood estimates, which do not strictly hold in many practical situations. It is necessary to have a lot of replications. In the most typical situations, the number of observations to number of parameters ratio must be at least 15 to completely rely on these statistics. Otherwise, one has to examine the behavior of these statistics case by case using Monte Carlo techniques.

The same example data sets that were previously analyzed by ADDALS and WADDALS were reanalyzed by MAXADD, assuming that the data were directional rank orders. Two analyses were performed—one under the assumption that the rankings were made from the smallest to the largest and the other under the assumption that the rankings were done in the reverse direction. (In a practical situation, one usually does only one analysis, whichever an experimental paradigm dictates.) Results are presented in Figure 13–10. The two analyses give similar results; not only are the values of the likelihood function quite similar, but estimates of the model parameters are also similar. These estimates also agree well with those obtained by WADDALS.

An example of some use of statistical inference features of MAXADD will be given in the application section.

Figure 13–10. Summary of MAXADD-4 Analyses of the Data for WADDALS

(I) Assuming that the rankings are performed from the smallest to the largest:

$LL = -13.6$
$n_\pi = 5$
$AIC = 37.3$

Additive effects (Normalized)

$$\alpha = \begin{pmatrix} -.520 \\ -.879 \\ 1.399 \end{pmatrix} \quad \beta = \begin{pmatrix} -1.275 \\ .108 \\ 1.167 \end{pmatrix}$$

Weights (Normalized)

	w_A	w_B
k=1	.789	.615
k=2	.579	.816

(II) Assuming that the rankings are performed from the largest to the smallest:

$LL = -13.9$
$n_\pi = 5$
$AIC = 37.7$

Additive effects (Normalized)

$$\alpha = \begin{pmatrix} -.220 \\ -1.100 \\ 1.320 \end{pmatrix} \quad \beta = \begin{pmatrix} -1.287 \\ .137 \\ 1.151 \end{pmatrix}$$

Weights (Normalized)

	w_A	w_B
k=1	.892	.452
k=2	.634	.773

EXAMPLES OF APPLICATION

In this section, we discuss some applications. Some of these examples have been presented elsewhere; others are only provisional, in that no real data have yet been analyzed, but they are included here to demonstrate possible uses of WADDALS or MAXADD. They also indicate how the two procedures can be effectively applied to actual data analysis situations. We focus on the WAM here, but the reader can find interesting applications of the SAM in Anderson (1981), Cliff (1959), Green and Rao (1971), Johnson (1974), and Wallsten (1976). Green and Srinivasan (1978) give an excellent overview of various current issues on the additive conjoint measurement.

Psychophysical Data

Portions of this example have been presented elsewhere (Takane et al. 1980; Takane 1982). In fact, it served as an original motivating factor for developing the WADDALS procedure, the first computer program ever developed for the WAM.

Some developmental psychologists believe that the way largeness of rectangles is judged changes with the subject's age. For

example, younger children tend to put more emphasis on the height than the width of rectangles when they make perceived largeness judgments. This tendency decreases as they get older (Kempler 1971). Let us assume (although this assumption itself is susceptible to an empirical test) that children's perception of the rectangle area is obtained by simple addition of contributions from the height and the width. That is, it is representable by the SAM for each child or age group.

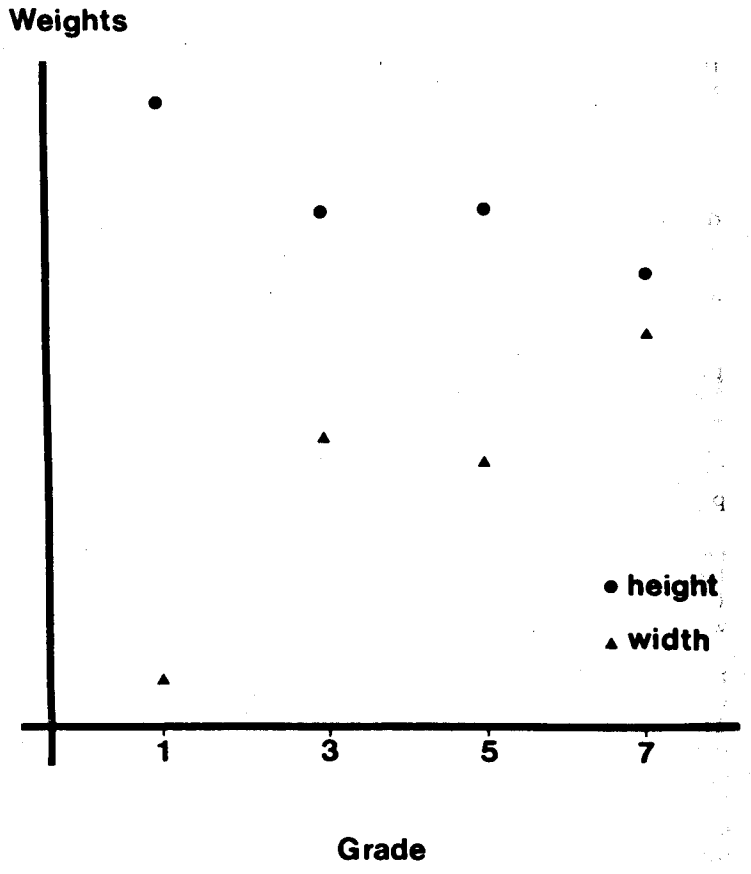
The developmental change described above may clearly be described by the WAM. Let w_{kH} and w_{kW} denote the differential weights put on the height and the width, respectively, in the kth age group. It is expected that $w_{kH} > w_{kW}$ initially and that w_{kH} decreases while w_{kW} increases as k increases, until w_{kH} becomes nearly equal to w_{kW}.

In order to test this hypothesis, Kempler (1971) constructed a set of 100 rectangles by factorially combining 10 height levels and 10 width levels, each ranging from 10 inches to 14.5 inches in half-inch intervals. Sixteen to 25 children in each of four age groups (1st-, 3rd-, 5th-, and 7th-graders) judged each of the 100 rectangles "large" or "small." (This is a two-category judgment.) For each age group, the number of times each rectangle is judged large (that is, the number of children who make this judgment) was counted and used as an ordinal measure of the perceived largeness and additivity analyses were applied.

Since the WAM presupposes a representation by the SAM for each age group, the simple additivity analysis was conducted first. In all cases, satisfactory stress values were obtained ($< .2$). Since the estimates of the additive effects in the SAM were found roughly proportional, the weighted additivity analysis was performed by WADDALS. As expected, the weights attached to the height tended to decrease, while those for the width increased rather consistently with age level (see Figure 13–11). The stress value obtained from WADDALS (.190) was not very different from the joint stress value obtained from the SAM (.172) applied separately to each age group. Thus, for Kempler's data, the WAM was indeed appropriate. Takane et al. (1980) discuss other possible models for Kempler's data and how they can be dismissed on the grounds of model comparisons. They also report extensive Monte Carlo studies on stress values to be used in evaluating the goodness-of-fit of the WAM.

The above analysis used group data, ignoring possible individual differences within each group. Thus, it did not reveal a true source of the group differences found. To do so, individual data have to be analyzed. Kempler's individual data (original two-category judgments) were thus analyzed separately for each age group by MAXADD using both WAM and SAM. The WAM was consistently found to fit the data better, implying that there were significant differences in the weights *within* each group. Figure 13–12 shows the estimated subject weights for each of the four age groups. There are indeed marked individual differences in the weights. What is more interesting, however, is that there seem to be some systematic differences in the weight structures across the groups. The plots of the weight estimates tend to converge toward the dotted lines (the 45° lines between horizontal and vertical axes) as the age level goes up. For example, in

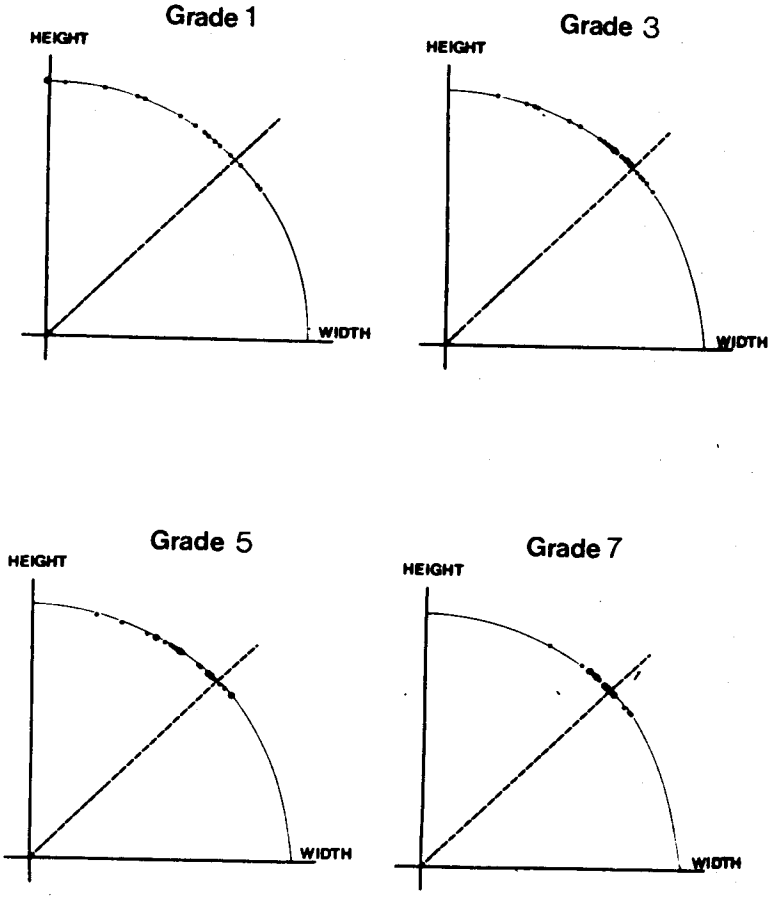
Figure 13–11. Individual Differences Weights Obtained from WADDALS Analyses of Group Data



grade 1, several children put disproportionately large weights on the height. (There are two children who base their largeness judgments almost exclusively on the height factor.) As the age level goes up, the number of those extreme children decreases. The children themselves also tend to be less extreme. In grade 7, a majority of children put approximately equal emphasis on the two factors. Thus, the group differences found in the previous analysis seem largely due to the compositional differences among the groups. That is, each group consists of heterogeneous subjects, but the degree of heterogeneity within each group clearly decreases with age.

In the analysis of the group data, the additive effects were found to be constant across the groups. In fact, this was the basis for applying the WAM to all age groups simultaneously. Can they still be assumed constant across the different age groups with the individual data? (Remember that in the above

Figure 13–12. Individual Differences Weights Obtained from MAXADD Analyses of Individual Data



analysis the WAM was applied separately to the four age groups, which tacitly assumed that the additive effects were constant across the subjects within the groups.) In order to answer this question, the WAM was applied to the entire set of individual data and the goodness-of-fit of this model was compared with that of the WAM applied separately to each group. The AIC from the joint analysis was 49.2 (+ constant), while the joint AIC from the separate analyses was 51.6 (+ constant). Thus, it was concluded that the WAM fitted to the entire set of individual data was the better-fitting model. The additive effects can indeed be assumed constant across the age groups.

Similarity and Dissimilarity Data

Medin and Schaffer (1978) have proposed a model called the *cue context model* for classification learning. In their model, simi-

larity between stimuli i and j is defined as:

$$s_{ij} = \prod_{t=1}^{T} X_{ijt} \quad , \tag{13–41}$$

where s_{ij} is the similarity between stimuli i and j, and

$$X_{ijt} = \begin{cases} 1, & \text{if stimuli } i \text{ and } j \text{ share feature } t, \\ \Theta_t, & (0 < \Theta_t \leq 1) \text{ otherwise.} \end{cases}$$

The X_{ijt} thus denotes the featurewise similarity between stimuli i and j. The overall similarity (s_{ij}) is defined as the product of the featurewise similarities. Since $0 < \Theta_t \leq 1$ for $t = 1, \ldots, T$, we have $0 < s_{ij} \leq 1$. Let

$$d_{ij} = -\log s_{ij} = -\sum_{t=1}^{T} \ln X_{ijt} \quad . \tag{13–42}$$

The d_{ij} represents a dissimilarity between i and j. (It is an inverse monotonic transformation of s_{ij}.) Since part-dissimilarity functions (Green and Srinivasan 1978), which are

$$d_{ijt} = -\ln X_{ijt} = \begin{cases} 0, & \text{if stimuli } i \text{ and } j \text{ share feature } t, \\ -\ln \Theta_t, & \text{otherwise,} \end{cases}$$

are nonnegative ($-\ln \Theta_t \geq 0$), the overall dissimilarity (d_{ij}) is also nonnegative. Furthermore, it can be proven that:

$$d_{ii} = 0 \text{ (Minimality)},$$

$$d_{ij} = d_{ji} \text{ (Symmetry)},$$

and

$$d_{ij} + d_{jk} \geq d_{ik} \text{ (Triangle Inequality)},$$

so that d_{ij} in (13–42) can formally be considered as the distance between stimuli i and j. Define

$$g_{ijt} = \begin{cases} 0, & \text{if stimuli } i \text{ and } j \text{ share feature } t, \\ 1, & \text{otherwise,} \end{cases}$$

and

$$\Theta_t^* = -\ln \Theta_t \quad .$$

Then $d_{ijt} = g_{ijt}\,\Theta_t^*$, and therefore:

$$d_{ij} = \sum_{t=1}^{T} d_{ijt} = \sum_{t=1}^{T} g_{ijt}\,\Theta_t^* \ . \qquad (13\text{–}43)$$

This indicates that d_{ij}, which is a kind of distance as defined in (13–42), is a special case of the SAM.

Note that if we define $d_{ij}^{p} = -\ln s_{ij}$ for any p (≥ 1) in (13–42), d_{ij} can be regarded as the Minkowski p-metric. (When $p = 2$, we obtain the Euclidean distance; when $p = 1$, as in the above discussion, we have the city-block distance.) We also define $d_{ijt}^{p} = -\ln X_{ijt}$ in this case. That p can be any value not less than unity reflects the fact that there is only one interval defined for each dimension.

Suppose now that we may allow individual differences in s_{ij} in the following manner:

$$s_{kij} = \prod_{t=1}^{T} X_{ijt}^{w_{kt}} \ , \qquad (13\text{–}44)$$

where s_{kij} is the similarity between stimuli i and j for individual k, and w_{kt} is the weight attached to feature t by individual k. Let

$$d_{kij}^{p} = -\ln s_{kij} = -\sum_{t=1}^{T} w_{kt} \ln X_{ijt} \quad \text{for } (p \geq 1).$$

Again, d_{kij} can be proven to be a distance function within each k. We obtain, analogous to (13–43),

$$d_{kij}^{p} = \sum_{t} g_{ijt}\,\Theta_t^* w_{kt} \ , \qquad (13\text{–}45)$$

which is a special case of the WAM. This is also a special case of (13–22), in which

$$g_{ijt}\,\Theta_t^* = |x_{it} - x_{jt}|^{p} \ .$$

Thus, sets of similarity or dissimilarity data for stimuli characterized by a set of known features may be analyzed using either WADDALS or MAXADD, depending on what assumptions we are willing to make on our data.

Model (13–45), however, reduces to an even simpler form by redefining $\Theta_{kt}^* = \Theta_t^* w_{kt}$. We then obtain:

$$d_{kij}^{p} = \sum_{t} g_{ijt}\,\Theta_{kt}^* \ , \qquad (13\text{–}46)$$

which amounts to applying the SAM separately to each k. This is because there are only two levels in each feature (present or absent). In this case, the WAM and the SAM are completely equivalent. For the same reason a correct value of the Minkowski power cannot be identified. In order to empirically distinguish the WAM from the SAM, and also between different values of p, we must have at least three levels in each feature (factor, dimension). When there are three levels—say i, q, and s—then there are three intervals defined—(i,q), (q,s), and (i,s). Contributions of these intervals should satisfy a special relationship under the Minkowski power distance model:

$$\delta_{iq}^{1/p} + \delta_{qr}^{1/p} = \delta_{is}^{1/p} \quad , \tag{13–47}$$

where δ_{iq} indicates the effect of the difference between two levels, i and q. (Under the Minkowski hypothesis, this is indeed the pth power of the length of an interval bounded by i and j.) If we use (13–47) as constraints while systematically varying the value of p, we should be able to identify the best-fitting p value. Takane (1982) reports this type of study with the SAM.

It may be noted in passing that the multiplicative rule in (13–41) in defining the overall similarity between two stimuli is quite similar to the multiplicative competitive interaction (MCI) model (Nakanishi and Cooper 1974) for preference choice data, to which we now turn.

Preference Data

Applications of the WAM to preference data are rather scarce, due, perhaps, in part to the fact that the WAM postulates an identical ideal point (Coombs [1964] 1976) along each stimulus attribute across individuals. In situations in which the preference functions are single-peaked, this is hard to justify; in many cases, there are individual differences in the ideal point. However, just as the ideal point model reduces to the vector preference model when the ideal points are located infinitely far away from the origin (Carroll 1972), the additive model in this limiting case posits monotonic preference functions for the additive effects. Just as in a number of situations the vector preference model is appropriate (Green and Srinivasan 1978), there are a lot of situations in which the preference functions (often called *part-worth functions* in the marketing research literature; see, for example, Green and Srinivasan 1978) can be assumed monotonic and common across individuals, thus making the WAM appropriate. The WAM in this case can be thought of as a special type of the vector preference model, which involves quantifications of attributes as well as weighting.

Phipps and Carter (1982) applied the WAM to neighborhood preference data of homeowners using lot size, landscaping, and neighboring housing as stimulus attributes. Along each of the three attributes, the preference function was found to be monotonic. The individual difference weights obtained from the

weighted additivity analysis were then related to various demographic variables of the individuals, such as education. C. P. Whaley (personal communication), of Bell Northern Research, also reports an application of WADDALS in marketing research called *tradeoff analysis* (see Johnson 1974). Unfortunately, details of the study cannot be presented, due to its proprietary nature. Nonetheless, in this study, there were five factors to be considered, but judgments were taken only for two factors at a time. From these incomplete data, WADDALS could successfully obtain estimates of parameters in the WAM.

Several preference models can be considered as special cases of the WAM. We will discuss only two of them, the *multiplicative competitive interaction* (MCI) *model* (Nakanishi and Cooper 1974) and the *additive-difference model* (Tversky 1969) *for preference data*.

The MCI model posits that the overall utility of stimulus r (y_r) is a powered multiplicative function of stimulus attributes (X_{rt}). That is,

$$y_r = \prod_{t=1}^{T} X_{rt}^{w_t} , \tag{13–48}$$

which can be readily extended to

$$y_{kr} = \prod_{t=1}^{T} X_{rt}^{w_{kt}} , \tag{13–49}$$

where w_{kt} is an individual-difference weight. Taking the log of (13–49), we obtain:

$$\log y_{kr} = \sum_{t=1}^{T} w_{kt} \log (X_{rt}) , \tag{13–50}$$

which is linear with respect to log (X_{rt}). Suppose now that log (X_{rt}) is unknown and that it should be quantified. Then

$$\log (X_{rt}) = \sum_{q_t=1}^{Q_t} h_{rq_t} \alpha_{q_t} .$$

Putting this into (13–50), we obtain:

$$\log y_k = \sum_{t=1}^{T} w_{kt} \sum_{q_t=1}^{Q_t} h_{rq_t} \alpha_{q_t} , \tag{13–51}$$

which is precisely the WAM (see [13–3]). In (13–51), we may just observe an ordinal measure of the dependent variable rather than log y_{kr} itself (such as y_{kr}). Then it must be monotonically transformed before it can be fitted by the WAM.

We now turn to the additive-difference model. There are actually two versions of the model: (a) a symmetric version designed for dissimilarity data (Tversky and Krantz 1970); and (b) a skew-symmetric version (Tversky 1969) designed for dominance data. It is this latter version that we discuss here for preference data.

When we compare two or more stimuli in terms of preference, we may first evaluate overall stimulus preferences by combining various stimulus attributes and then compare them across the stimuli to arrive at a relative preference judgment, that is, how much we like one stimulus over the other. The additive model is appropriate if the stimulus attributes are combined in an additive fashion. Alternatively, we may first compare the stimuli within attributes and then, based on the attribute by attribute comparisons, form an overall preference judgment. In this case, an additive representation no longer holds except under a specialized condition (Tversky 1969).

Let us suppose for simplicity that the attributewise comparison process is subtractive and that the subsequent integration process is weighted additive. Then we obtain the following model:

$$y_{kijuv} = w_{kA}\psi_A(\alpha_i - \alpha_u) + w_{kB}\psi_B(\beta_j - \beta_v), \tag{13–52}$$

where ψ's are some monotonic skew-symmetric functions ($\psi(-x) = -\psi(x)$), and y_{kijuv} is the relative preference of stimulus (i,j) (the stimulus representing the combined effect of the ith level of factor A and the jth level of factor B) in comparison with stimulus (u,v) by individual k. The above model is called the *weighted additive-difference model for preference data* (Takane 1982). The additive-difference model reduces to the difference between two additive models when ψ's are linear (in which case ψ's can further be assumed to be identity functions), since we obtain:

$$\begin{aligned} y_{kijuv} &= w_{kA}(\alpha_i - \alpha_u) + w_{kB}(\beta_j - \beta_v) \qquad (13–53)\\ &= (w_{kA}\alpha_i + w_{kB}\beta_j) - (w_{kA}\alpha_u + w_{kB}\beta_v)\\ &= y_{kij} - y_{kuv}. \end{aligned}$$

Thus, the WAM is a special case of the weighted additive-difference model in which ψ's are linear. In this special case, the two models are completely equivalent; consequently, the two comparison processes, stimulus-by-stimulus and attribute-by-attribute, are not empirically distinguishable.

In general, (13–52) is not reducible to (13–53). However, in some sense, the weighted additive-difference model can also be considered as a special case of the weighted additive model. We redefine differences between levels of additive factors (that is, $\alpha_i - \alpha_u$, $\beta_j - \beta_v$, and so on) as new levels of additive factors.

Then we can directly estimate their effects in the context of the additive model. Let $\alpha_{iu} = \psi_A(\alpha_i - \alpha_u)$ and $\beta_{jv} = \psi_B(\beta_j - \beta_v)$. Then (13–52) can be rewritten as:

$$y_{kijuv} = w_{kA}\alpha_{iu} + w_{kB}\beta_{jv}, \tag{13–54}$$

where α_{iu} and β_{jv} represent the additive effects of newly defined levels of additive factors. We may require

$$\alpha_{iu} = -\alpha_{iu} \qquad \text{for all } i \text{ and } u \tag{13–55}$$

(the same for β), which follows from the skew-symmetry of ψ_A. It follows from (13–55) that:

$$\alpha_{ii} = 0 \qquad \text{for all } i. \tag{13–56}$$

This implies that if two stimuli share the same level of an attribute, the contribution of that attribute is zero in the comparison process. In order to preserve the monotonicity of ψ's, we may impose order restrictions, such as:

$$\alpha_{iu} \leq \alpha_{is} \qquad \text{and} \qquad \alpha_{us} \leq \alpha_{is},$$

where level u is between i and s. In any case, the goodness-of-fit of (13–54) can be directly compared with that of (13–53). This comparison should provide important insight into the nature of stimulus comparison processes.

Note that restriction (13–55) leads, strictly speaking, to a nonadditive model (though it is still linear), since we obtain model predictions such as:

$$y_{kijiv} = -w_{kA}\alpha_{iu} + w_{kB}\beta_{jv},$$

$$y_{kivuj} = w_{kA}\alpha_{iu} - w_{kB}\beta_{jv},$$

and so forth. Unfortunately, the nonadditive model cannot be directly fitted by WADDALS or MAXADD. Necessary modifications should not be too difficult, however. MAXLIN, for the general weighted linear model, is now under development, which should be able to handle this kind of situation.

Analysis of Contingency Tables

Tables of joint frequencies of responses to two or more multicategory items are called *contingency tables*. The log-linear model (Bishop et al. 1975) has been quite popular for analysis of the contingency tables.

Let f_{ij} denote the observed frequency of joint occurrences of the ith category (level) of item (factor) A and the jth category of

item B. (For simplicity, we only deal with a two-item situation.) Then, the log-linear model states that:

$$p_{ij} \equiv f_{ij}/N = \alpha_i^* \beta_j^* \gamma_{ij}^* \quad , \tag{13–57}$$

where $N = \sum_i \sum_j f_{ij}$ (the total number of observations),

α_i^* and β_j^* are main effects of the ith category of item A and the jth category of item B, respectively,

and γ_{ij}^* is an *interaction* effect between i and j.

There are too many parameters in the above model, so that constraints such as

$$\prod_i \alpha_i^* = \prod_j \beta_j^* = \prod_i \gamma_{ij}^* = \prod_j \gamma_{ij}^* = 1$$

are imposed to identify the model parameters. Taking the log of (13–57), we obtain:

$$\log p_{ij} = \alpha_i + \beta_j + \gamma_{ij} \quad , \tag{13–58}$$

where

$$\alpha_i = \log \alpha_i^* \quad ,$$

$$\beta_j = \log \beta_j^* \quad ,$$

and

$$\gamma_{ij} = \log \gamma_{ij}^* \quad ,$$

with

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0.$$

If we assume that there is no interaction effect in (13–58), we obtain the SAM,

$$\log p_{ij} = \alpha_i + \beta_j \quad , \tag{13–59}$$

which may equivalently be written as:

$$p_{ij} = \exp(\alpha_i + \beta_j) \quad . \tag{13–60}$$

This latter form of the model suggests that it is a special case of Luce's (1959) choice model,

$$p_{ij} = \exp(\alpha_i + \beta_j) / \sum_{i'} \sum_{j'} \exp(\alpha_{i'} + \beta_{j'}) , \qquad (13\text{–}61)$$

with the denominator scaled to be unity. Model (13–61) can be obtained alternatively by assuming that the cell (i,j) is *chosen* with the probability proportional to its response strength relative to those of other cells. That is, if we take the SAM $(\alpha_i + \beta_j)$ for the representation of (i,j) cell and use the model of first choice, similar to the one used in the directional ranking method, we obtain (13–61). Model (13–61) also has some resemblance to the Rasch model (Andersen 1980) in mental testing situations.

The extension of (13–59) or (13–60) to the WAM is rather straightforward; we obtain:

$$p_{kij} = \exp(w_{kA}\alpha_i + w_{kB}\beta_j) . \qquad (13\text{–}62)$$

The WAM in this case implies no interaction between item A and item B at all k's and hence no three-way interaction. It does imply a special type of two-way interaction between i and $k(w_{kA}\alpha_i)$ and between j and $k(w_{kB}\beta_j)$.

CONCLUSION

In this chapter, we have discussed various aspects of the WAM. We have seen that a wide variety of existing models can be thought of as special cases of the WAM. Although examples of its application are not yet sufficient, we believe that they will quickly grow in number in the near future. We hope that this discussion will serve to disseminate the basic idea behind the model and to encourage more applied work in the social sciences.

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19:716–23.

Andersen, E. B. 1980. *Discrete statistical models with social science applications*. Amsterdam: North-Holland.

Anderson, N. H. 1981. *Foundations of information integration theory*. New York: Academic Press.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland. 1975. *Discrete multivariate analysis: Theory and practice*. Cambridge, Mass.: The MIT Press.

Bock, R. D. 1975. *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

Carroll, J. D. 1972. Individual differences and multidimensional scaling. In *Multidimensional scaling: Theory and applications in the behavioral sciences*, vol. 1, ed. R. N. Shepard, A. K. Romney, and S. Nerlove. New York: Seminar Press.

———. 1973. Categorical conjoint measurement. In *Multiattribute decisions in marketing: A measurement approach*, ed. P. E. Green and Y. Wind. Hinsdale, Ill.: Dryden Press.

Carroll, J. D., and J. J. Chang. 1970. Analysis of individual

differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika* 35:283–319.

Cliff, N. 1959. Adverbs as multipliers. *Psychological Review* 66:27–44.

Coombs, C. H. [1964] 1976. *A theory of data.* [New York: Wiley] Republished, Ann Arbor, Mich.: Mathesis Press.

Cronbach, L. J., and G. Gleser. 1953. Assessing similarity between profiles. *Psychological Bulletin* 6:456–73.

de Leeuw, J., F. W. Young, and Y. Takane. 1976. Additive structure in qualitative data: An alternating least-squares method with optimal scaling features. *Psychometrika* 41: 471–503.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *The Journal of the Royal Statistical Society (Series B)* 39:1–38.

De Sarbo, W. S., J. D. Carroll, D. R. Lehmann, and J. O'Shaughnessy. Forthcoming. Three-way multivariate conjoint analysis, *Marketing Science.*

Draper, N. R., and H. Smith. 1981. *Applied regression analysis.* 2d ed. New York: Wiley.

Goodman, L. A. 1978. *Analyzing qualitative/categorical data: Log-linear models and latent-structure analysis.* Cambridge: Abt Books.

Green, P. E., and V. Rao. 1971. Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research* 8:355–63.

Green, P. E., and V. Srinivasan. 1978. Conjoint analysis in consumer research: Issues and outlook. *The Journal of Consumer Research* 5:103–23.

Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika* 33:469–506.

Hayashi, C. 1952. On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics* 2:69–98.

Johnson, R. M. 1974. Trade-off analysis of consumer values. *Journal of Marketing Research* 11:121–27.

Kempler, B. 1971. Stimulus correlates of area judgments: A psychophysical developmental study. *Developmental Psychology* 4:158–63.

Krantz, D. H., R. D. Luce, P. Suppes, and A. Tversky. 1971. *Foundations of measurement,* vol. 1. New York: Academic Press.

Kruskal, J. B. 1964a. Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika* 29:1–27.

————. 1964b. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29:115–29.

————. 1965. Analysis of factorial experiments by estimating monotone transformations of data. *Journal of the Royal Statistical Society (Series B)* 27:251–65.

————. 1971. Monotone regression: Continuity and differentiability properties. *Psychometrika* 36:57–62.

Kruskal, J. B., and R. N. Shepard. 1974. A nonmetric variety of linear factor analysis. *Psychometrika* 39:123–57.

Lingoes, J. C. 1973. *The Guttman-Lingoes nonmetric program series.* Ann Arbor, Mich.: Mathesis Press.

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

Luce, R. D., and J. W. Tukey. 1964. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology* 1:1–27.

Marley, A. A. J. 1970. Additive conjoint measurement with respect to a pair of orderings. *Philosophy of Science* 37: 215–22.

McDonald, R. P. 1968. A unified treatment of the weighting problem. *Psychometrika* 33:351–81.

Medin, D. L., and M. M. Schaffer. 1978. Context theory of classification learning. *Psychological Review* 85:207–38.

Nakanishi, M., and L. G. Cooper. 1974. Parameter estimation for a multiplicative competitive interaction model—Least-squares approach. *Journal of Marketing Research* 11:303–11.

Phipps, A. G., and J. E. Carter. 1982. Neighbourhood preferences of homeowners living in inner city areas affected by apartment redevelopment. Manuscript, University of Saskatchewan.

Ramsay, J. O. 1977. Monotonic weighted power transformations to additivity. *Psychometrika* 42:83–109.

———. 1978. Confidence regions for multidimensional scaling analysis. *Psychometrika* 43:145–160.

Rao, V. R. 1977. Conjoint measurement in marketing analysis. In *Multivariate methods for market and survey data,* ed. J. N. Sheth. Chicago: American Marketing Association.

Roskam, E. E. 1968. *Metric analysis of ordinal data in psychology.* Voorschoten, Holland: VAM.

Sayeki, Y. 1972. Allocation of importance: An axiom system. *Journal of Mathematical Psychology* 9:55–65.

Shepard, R. N. 1972. A taxonomy of some principal types of data and of multidimensional methods for their analysis. In *Multidimensional scaling,* vol. 1, ed. R. N. Shepard et al. New York: Seminar Press.

Srinivasan, V., and A. D. Shocker. 1973. Estimating the weights for multiple attributes in a composite criterion using pairwise judgments. *Psychometrika* 38:473–93.

Takane, Y. 1978. A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—Theory and evaluations. *Japanese Psychological Research* 20:7–17, 105–14.

———. 1981. Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika* 46:9–28.

———. 1982. Maximum likelihood additivity analysis. *Psychometrika* 47:225–41.

Takane, Y., and J. D. Carroll. 1981. Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities. *Psychometrika* 46:389–405.

Takane, Y., F. W. Young, and J. de Leeuw. 1980. An individual differences additive model: An alternating least-squares method with optimal scaling features. *Psychometrika* 45:183–

209.
Tversky, A. 1969. Intransitivity of preferences. *Psychological Review* 73:31–48.
Tversky, A., and D. H. Krantz. 1970. The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology* 7:572–96.
Wallsten, T. S. 1976. Using conjoint-measurement models to investigate a theory about probabilistic information processing. *Journal of Mathematical Psychology* 14:144–85.
Winsberg, S., and J. O. Ramsay. 1980. Monotonic transformations to additivity using splines. *Biometrika* 67:669–74.
Young, F. W. 1981. Quantitative analysis of qualitative data. *Psychometrika* 46:357–88.
Young, F. W., J. de Leeuw, and Y. Takane. 1976. Regression with qualitative and quantitative variables: An alternating least-squares method with optimal scaling features. *Psychometrika* 41:505–29.