PSYCHOMETRIKA — VOL. 45, NO. 1.
JUNE, 1980

AN INDIVIDUAL DIFFERENCES ADDITIVE MODEL: AN ALTERATING LEAST SQUARES METHOD WITH OPTIMAL SCALING FEATURES

YOSHIO TAKANE

MCGILL UNIVERSITY

FORREST W. YOUNG

UNIVERSITY OF NORTH CAROLINA

JAN DE LEEUW

UNIVERSITY OF LEIDEN

An individual differences additive model is discussed which represents individual differences in additivity by differential weighting of additive factors. A procedure for estimating the model parameters for various data measurement characteristics is developed. The procedure is evaluated using both Monte Carlo and real data. The method is found to be very useful in describing certain types of developmental change in cognitive structure, as well as being numerically robust and efficient.

Key words: additivity analysis, individual differences, optimal scaling, conjoint scaling.

1. Introduction

The analysis of additivity centers on the conditions under which a dependent measure can be represented as an additive function of independent variables. Multidimensional scaling is a common type of additivity analysis since the Minkowski distance model is an additive conjoint model [Krantz, Luce, Suppes & Tversky, 1971]. Similarly, the weighted distance model to represent individual differences [Carroll & Chang, 1970] is a kind of weighted additive conjoint measurement whose axiomatic foundation has been given by Sayeki [1972]. In the weighted additive conjoint model the individual differences are represented by differential weighting of additive factors, while the additive factors themselves are assumed to have identical forms across individuals. In this paper we describe an alternating least squares (ALS) procedure to fit a weighted additive model. Our model is an extension of the simple additive model that is analogous to the individual differences extension of the distance model. The fitting procedure, called WADDALS (*Weighted Additive Model by Alternating Least Squares*), generalizes our previous developments on linear models, ADDALS [de Leeuw, Young & Takane, 1976] and MORALS [Young, de Leeuw & Takane, 1976b] by estimating individual differences weights in a fashion similar to that in ALSCAL [Takane, Young & de Leeuw, 1977].

The work reported here was partly supported by Grant A6394 to the first author by the Natural Sciences and Engineering Research Council of Canada.

Requests for reprints should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Docteur Penfield Avenue, Montreal, Quebec, H3A 1B1.

0033-3123/80/0600-2677$00.75/0

Perhaps an example would help clarify the situation for which the present model is most suitable. The physical area of a rectangle, which is defined by

$$\text{area} = \text{height} \times \text{width},$$

is a simple example of additive conjoint measurement. This can be easily seen by applying the log transformation to both sides of the above equation. We then have

$$\log(\text{area}) = \log(\text{height}) + \log(\text{width})$$

which is a special case of

$$f(\text{area}) = f_H(\text{height}) + f_W(\text{width}), \tag{1}$$

where f, f_H and f_W are monotonically increasing functions of physical area, height and width of a rectangle, respectively. This latter equation is a general expression of additive conjoint measurement for two factors. An interesting question (in terms of multidimensional psychophysics) is whether the subjective area of rectangles has an additive representation implied by (1). We have previously developed a procedure called ADDALS [de Leeuw, et al, 1976; see also Kruskal, 1965] which is specifically designed to deal with this type of simple additivity problem.

Suppose that the additive representation (1) holds for the subjective area of rectangles (within a reasonable amount of error) for each individual subject. The next question to be posed might be how the individual's perceptions are related to each other. It might be that representations of the individual's perceptions (albeit additive) are not related in a simple way. In this case (1) should be applied separately to each individual's judgments. That is,

$$f_k(\text{area}) = f_{Hk}(\text{height}) + f_{Wk}(\text{width}), \tag{1'}$$

where f_k, f_{Hk} and f_{Wk} are unique for each individual (k), and no particular relationships are assumed among f_{Hk} or among f_{Wk} over different individuals. Alternatively, we may find a simple relationship among individuals' perceptual structures of the largeness of rectangles. For example, as is well-known, younger children tend to place more emphasis on height than on width of rectangles when they make subjective largeness judgments of rectangles. This observation tempts one to hypothesize that they put more weight on height of rectangles than adults do when they combine height and width into a global largeness judgment. This hypothesis can be written as

$$f_k(\text{area}) = w_{Hk} f_H(\text{height}) + w_{Wk} f_W(\text{width}), \tag{2}$$

where w_{Hk} and w_{Wk} represent the weights attached to the height and width dimensions of rectangles, respectively, by individual k. Model (2) assumes that the transformation processes (f_H and f_W) of height and width remain the same for all individuals, but that they are differentially weighted (w_{Hk} and w_{Wk}) to produce different area judgments (f_K) by different individuals. Whether this supposition is adequate or not is an empirical question, and must be subjected to empirical testing. This can be done by comparing the goodness of fit of (2) against that of (1'). In this paper we develop a fitting procedure for (2), which enables us to make such a comparison.

In the next section we discuss various characteristics of the weighted additive model with emphasis on its geometric features and on its relation to the simple additive model. The fitting procedure (the optimization criterion and an associated algorithm) will be presented in Section 3. The model and the performance of the estimation procedure will be evaluated in Section 4 by Monte Carlo techniques. Finally an application of the procedure to empirical data is described to demonstrate its use.

2. *The Model and Its Properties*

For the purpose of data analysis it is convenient to have a collective expression for model (2) for the entire set of stimuli employed in an experiment. To illustrate let us go back for a moment to the example in which an investigator is interested in finding out whether the individual differences in the perceived largeness of rectangles can be represented by the weighted additive model. In this case he might construct a set of rectangles (stimuli) by factorially combining different height and width levels, and obtain a set of largeness judgments over the set of stimuli from a sample of individuals. The set of observations obtained from an individual subject may be placed into a vector in a prescribed order. This vector is denoted by $\mathbf{o}_k$ for individual k. The set of rectangles, on the other hand, may be characterized by several design matrices, one for the height and one for the width factor. Each row of each design matrix has exactly one element equal to unity (and zeroes elsewhere), where the unity indicates that the stimulus corresponding to the row is associated with the level of the factor corresponding to the column. For example, if $\mathbf{G}_H$ is the design matrix for the height factor, then a one in the i^{th} row and j^{th} column means that the i^{th} rectangle has the j^{th} height level.

Given an n stimuli by n_s levels design matrix $\mathbf{G}_s$ for factor s (s may be either the height factor or the width factor in the above example), we can express the additive contribution of the factor by

$$\mathbf{x}_s = \mathbf{G}_s\mathbf{a}_s, \tag{3}$$

where $\mathbf{a}_s$ is an n_s-element vector of additive effects of factor s. Each of the n elements of $\mathbf{x}_s$ represents an additive contribution from factor s for a stimulus associated with a certain level of that factor. Let w_{sk} denote the weight attached to factor s by individual k. Then a set of predictions from the weighted additive model can be written as

$$\hat{\mathbf{y}}_k = \mathbf{x}_H w_{Hk} + \mathbf{x}_W w_{Wk} \tag{4}$$

for individual k, where $\hat{\mathbf{y}}_k$ is a vector of predicted largeness of rectangles.

The vector of observations $\mathbf{o}_k$ is assumed to be related to the vector of model predictions $\hat{\mathbf{y}}_k$ through function ξ_k, though this relation is only approximate in the presence of random error. That is,

$$\xi_k(\mathbf{o}_k) = \mathbf{y}_k^* \cong \hat{\mathbf{y}}_k. \tag{5}$$

Here $\mathbf{y}_k^*$ is the vector of transformed values of observations and is sometimes called the vector of optimally scaled data. The symbol "$\cong$" indicates that quantities on both sides are approximately equal. The transformation ξ_k is to allow for observations with various measurement characteristics [Young, de Leeuw & Takane, 1979] such as scale levels of measurement. The same model may be fit to observations with various measurement characteristics by assuming various functional forms on ξ_k. For example, if observations are ordinal then the transformation must be an order-preserving function. The type of a transformation deemed appropriate under specific measurement characteristics is called the admissible transformation, within which the specific form of the transformation is sought based on some optimality consideration. That is, the functional form of ξ_k is determined in such a way that $\mathbf{y}_k^*$ in (5) agrees with $\hat{\mathbf{y}}_k$ in (4) as much as possible (in the sense defined in Section 3) under the constraints (e.g., monotonicity) imposed by specific measurement characteristics assumed of observations. Values of model parameters ($\mathbf{a}_H$, $\mathbf{a}_W$, w_{Hk} and w_{Wk}), on the other hand, are determined so that $\hat{\mathbf{y}}_k$ derived from the parameter values is as close to $\mathbf{y}_k^*$ as possible (based on the same optimality consideration as above) under the constraints imposed by the specific combination rule of those parameters.

Formal Statement of the Model

Let us formalize the above discussion in more general terms. Suppose that there are N individuals, each responding to a set of n stimuli with respect to some prescribed stimulus attribute. The stimuli are constructed by factorially combining levels of two or more externally identifiable stimulus attributes (called factors). Suppose we have n_s levels in factor s. The set of stimuli can be characterized by a set of n-by-n_s design matrices G_s where s extends over the set of factors (denoted by S) under study. We assume that measurements taken on the stimulus attribute can be represented by an additive combination of the effects due to factors (defining the set of stimuli). The effect due to factor s is denoted by an n_s-component vector $\mathbf{a}_s$ ($s \in S$). The additive effects are assumed common to all individuals, while those additive effects are assumed differentially weighted by individuals to give rise to different judgments over individuals. (Note that $\mathbf{a}_s$ does not have an index for an individual.) The differential weighting scheme of the additive effects is denoted by w_{sk} ($s \in S$; $k = 1, \cdots, N$), the weight attached to factor s by individual k.

We then formally state the weighted additive model as

$$\xi_k(\mathbf{o}_k) = \mathbf{y}_k^* \cong \underline{\hat{\mathbf{y}}_k} = \sum_{s \in S} \mathbf{G}_s \mathbf{a}_s w_{sk} \quad (k = 1, \cdots, N) \tag{6}$$

where ξ_k is a vector valued function defining the optimal scaling (transformation) of the observed data under various measurement restrictions, $\mathbf{o}_k$ is a vector of observed data, $\mathbf{y}_k^*$ is a vector of optimally transformed data, and $\hat{\mathbf{y}}_k$ is a vector of model predictions. (All the vectors defined above have n components.) The index k designates the k^{th} individual. The word individual, however, may refer to various entities such as group, occasion, experimental treatment, etc., so far as their effects can reasonably be presumed to be differential weightings of additive factors (rather than additive).

We assume, for the moment, that the experimental design is balanced without missing data and completely factorial, so that there are $N^* = N \times n$ observations in total where $n = \Pi_{s \in S} n_s$. (The unbalanced case will be discussed in the next section.) We further assume that the model predictions, $\hat{\mathbf{y}}_k$, and consequently the optimally scaled data, $\mathbf{y}_k^*$, are centered within each individual. This assumption reduces the weighted additive model to the simple additive model within each individual. More specifically, note that in the weighted additive model affine transformations of additive effects ($\tilde{\mathbf{a}}_s = c\mathbf{a}_s + d_s\mathbf{1}_{n_s}$ for some constants $c > 0$ and d_s for each $s \in S$ where $\mathbf{1}_{n_s}$ is an n_s-component vector of ones) incur an affine transformation of model predictions for each individual. That is

$$\tilde{\mathbf{y}}_k(\tilde{\mathbf{a}}_s) = c\hat{\mathbf{y}}_k + d_k\mathbf{1}_n \quad (k = 1, \cdots, N),$$

where $\tilde{\mathbf{y}}_k(\tilde{\mathbf{a}}_s)$ is a vector of model prediction derived from $\tilde{\mathbf{a}}_s$ and where $d_k = \Sigma_{s \in S} d_s w_{sk}$. However, since d_k is specific to individual k, affine transformations of $\mathbf{a}_s$ incur a nonaffine transformation for the set of $\hat{\mathbf{y}}_k$ over different individuals. Conversely, an affine transformation of individual $\hat{\mathbf{y}}_k$ preserves the set of $\mathbf{a}_s$ for $s \in S$. As in the simple additive model, we may optionally impose order restrictions on the elements of $\mathbf{a}_s$.

There are scale and sign indeterminacies inherent in the model (i.e., for an arbitrary $c \neq 0$, $\mathbf{a}_s^* = \mathbf{a}_s/c$ and $w_{sk}^* = cw_{sk}$ we have $\mathbf{a}_s^* w_{sk}^* = \mathbf{a}_s w_{sk}$). In order to eliminate these indeterminacies, we may impose, for identification purposes, the normalization restriction,

$$\frac{\mathbf{a}_s' \mathbf{G}_s' \mathbf{G}_s \mathbf{a}_s}{n} = \frac{\mathbf{a}_s' \mathbf{a}_s}{n_s} = 1, \tag{7}$$

for $s \in S$, and the sign restriction,

$$\max(a_{s_i}) > 0, \tag{8}$$

for $s \in S$ where a_{s_i} is the i^{th} element of $\mathbf{a}_s$. (If we explicitly constrain the individual differences weight w_{sk} to be nonnegative, the additional sign restriction is not necessary. Note, however, that the nonnegativity constraints on all w_{sk}'s are much stronger than (8).)

The weighted additive model defined above postulates additivity of the effects of the independent variables, $\mathbf{a}_s$ ($s \in S$), common to all individuals, but postulates that the differences among sets of data are due to different, but still additive, combinations of the common effects. Alternatively (though it amounts to the same thing) the model may be interpreted as postulating a multiplicative decomposition,

$$\mathbf{a}_{sk} = \mathbf{a}_s w_{sk} \quad (k = 1, \cdots, N)$$

for each $s \in S$ where $\mathbf{a}_{sk}$ is the vector of regression coefficients of $\mathbf{y}_k^*$ onto $\mathbf{G}_s$.

Geometric Features of the Model

The weighted additive model has a straightforward geometric interpretation which is illustrated in Figure 1. For explanatory convenience the simplest case in which there are only two individuals and only two factors, is shown in the figure. The vectors $\mathbf{y}_k^*$ and $\mathbf{y}_{k'}^*$ of optimally scaled data for individuals k and k', are orthogonally projected onto the space $\Omega(\mathbf{G}_A, \mathbf{G}_B)$ jointly spanned by the column vectors of the design matrices, $\mathbf{G}_A$ and $\mathbf{G}_B$, to obtain the vectors of model predictions, $\hat{\mathbf{y}}_k$ and $\hat{\mathbf{y}}_{k'}$, which are further decomposed into additive combinations of the projections of the two vectors on $\Omega(\mathbf{G}_A)$ along $\Omega(\mathbf{G}_B)$, and on $\Omega(\mathbf{G}_B)$ along $\Omega(\mathbf{G}_A)$, where $\Omega(\mathbf{G}_A)$ and $\Omega(\mathbf{G}_B)$ are the linear subspaces spanned by the column vectors of $\mathbf{G}_A$ and $\mathbf{G}_B$, respectively. The central characteristic of the model is that these projections point to precisely the same (or exactly the opposite) directions on each subspace, $\mathbf{G}_A\mathbf{a}_A \in \Omega(\mathbf{G}_A)$ and $\mathbf{G}_B\mathbf{a}_B \in \Omega(\mathbf{G}_B)$; only their lengths vary [proportionally to $w_{Ak}/w_{Ak'}$ for the two projections on $\Omega(\mathbf{G}_A)$ and $w_{Bk}/w_{Bk'}$ for the two projections on $\Omega(\mathbf{G}_B)$]. Of course, with fallible data the orientations of the projection vectors may not exactly coincide. However, if they are reasonably close, it will be meaningful to constrain them to coincide and to estimate the model parameters while assuming that the model is correct. This reduces the number of parameters to be estimated, and renders a more parsimonious description of certain aspects of individual differences in additivity than analyzing each individual's data separately by the simple additive model.

Note that the weighted additive model imposes a rather stringent set of assumptions upon the data structure. It should be clear from the preceding paragraph that perfect fit of the simple additive model to each and every individual's data set separately does not imply that the weighted additive model will perfectly fit the data as a whole. This is because the additive effects of each factor [the regression coefficients of $\mathbf{y}_k^*$ on $\Omega(\mathbf{G}_s)$] must be proportional to each other across individuals.

The Relation to the Simple Additive Model

There is a simple relationship between the weighted additive model and the simple additive model for fallible data. Figure 2 illustrates this relationship. If we orthogonally project vectors of optimally scaled data, $\mathbf{y}_k^*$ and $\mathbf{y}_{k'}^*$, onto the space spanned by $\mathbf{G}_s$, we obtain vectors of additive contributions from factor s, $\mathbf{G}_s\mathbf{a}_{sk}$ and $\mathbf{G}_s\mathbf{a}_{sk'}$, in the simple additive model. If we further project these vectors onto vector $\mathbf{G}_s\mathbf{a}_s$, which is constrained to be common to k and k' by the weighted additive model, we obtain vectors of additive contributions from factor s, $\mathbf{G}_s\mathbf{a}_s w_{sk}$ and $\mathbf{G}_s\mathbf{a}_s w_{sk'}$, in the weighted additive model. We have

$$\mathbf{G}_s\mathbf{a}_{sk} = \mathbf{G}_s(\mathbf{G}_s'\mathbf{G}_s)^{-1}\mathbf{G}_s'\mathbf{y}_k^*.$$

If we project this vector onto $\mathbf{G}_s\mathbf{a}_s$, the regression coefficient w_{sk} is given by

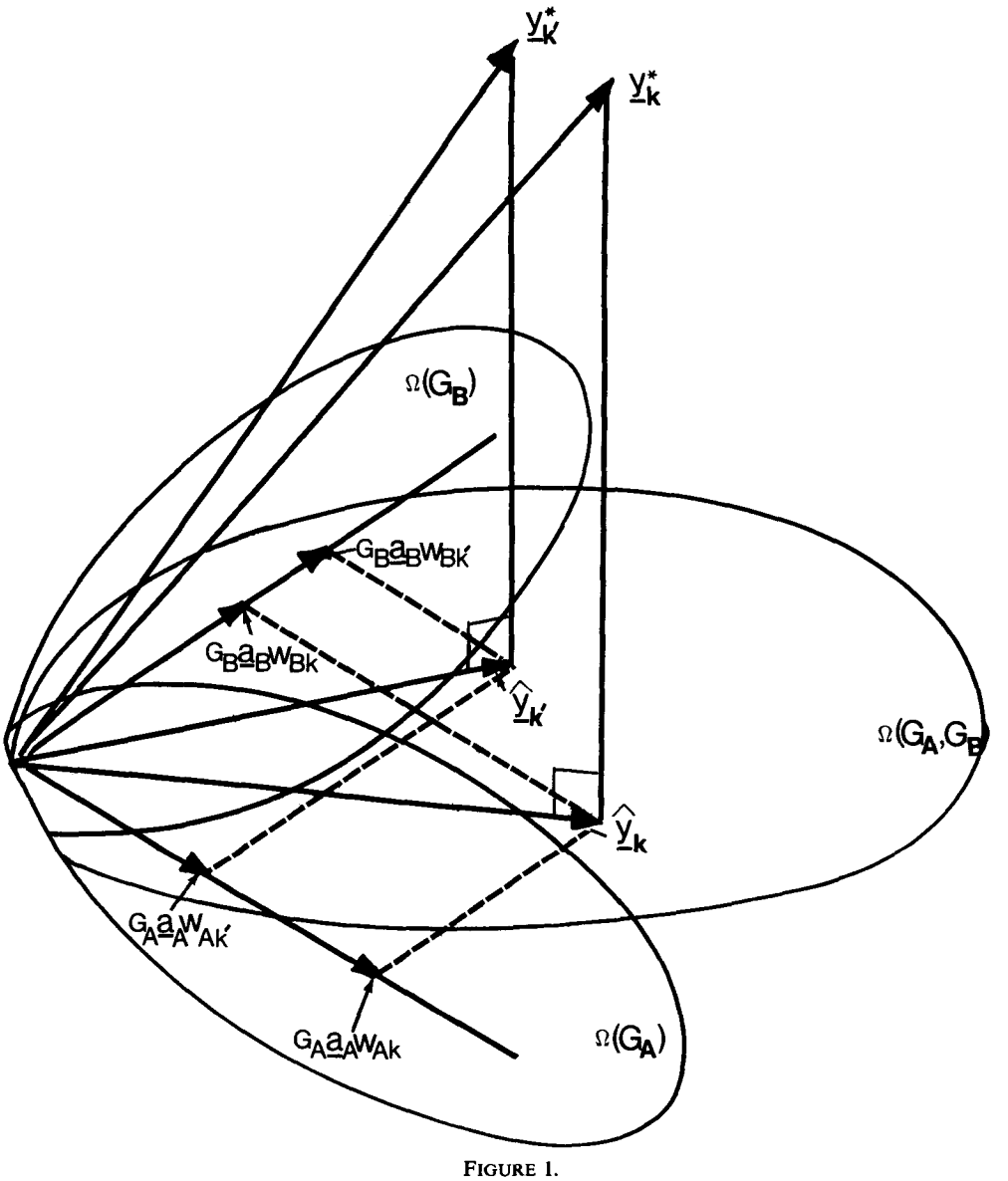

FIGURE 1.
A geometric interpretation of the weighted additive model.

$$(9) \qquad w_{sk} = (\mathbf{a}_s'\mathbf{G}_s'\mathbf{G}_s\mathbf{a}_s)^{-1}\mathbf{a}_s'\mathbf{G}_s'\mathbf{G}_s(\mathbf{G}_s'\mathbf{G}_s)^{-1}\mathbf{G}_s'\mathbf{y}_k^* = (\mathbf{a}_s'\mathbf{G}_s'\mathbf{G}_s\mathbf{a}_a)^{-1}\mathbf{a}_s'\mathbf{G}_s'\mathbf{y}_k^*,$$

which is equal to the regression coefficient obtained by directly projecting $\mathbf{y}_k^*$ onto $\mathbf{G}_s\mathbf{a}_s$.

Conversely, if w_{sk} is given by (9), $\mathbf{G}_s\mathbf{a}_s w_{sk}$ and $\mathbf{G}_s\mathbf{a}_{sk} - \mathbf{G}_s\mathbf{a}_s w_{sk}$ are shown to be orthogonal to each other:

$$\begin{aligned} & w_{sk}\mathbf{a}_s'\mathbf{G}_s'(\mathbf{G}_s\mathbf{a}_{sk} - \mathbf{G}_s\mathbf{a}_s w_{sk}) \\ &= w_{sk}\mathbf{a}_k'\mathbf{G}_s'\mathbf{G}_s(\mathbf{G}_s'\mathbf{G}_s)^{-1}\mathbf{G}_s'\mathbf{y}_k^* - w_{sk}^2\mathbf{a}_s'\mathbf{G}_s'\mathbf{G}_s\mathbf{a}_s \\ &= w_{sk}^2\mathbf{a}_s'\mathbf{G}_s'\mathbf{G}_s\mathbf{a}_s - w_{sk}^2\mathbf{a}_s'\mathbf{G}_s'\mathbf{G}_s\mathbf{a}_s = 0. \end{aligned}$$

We thus have

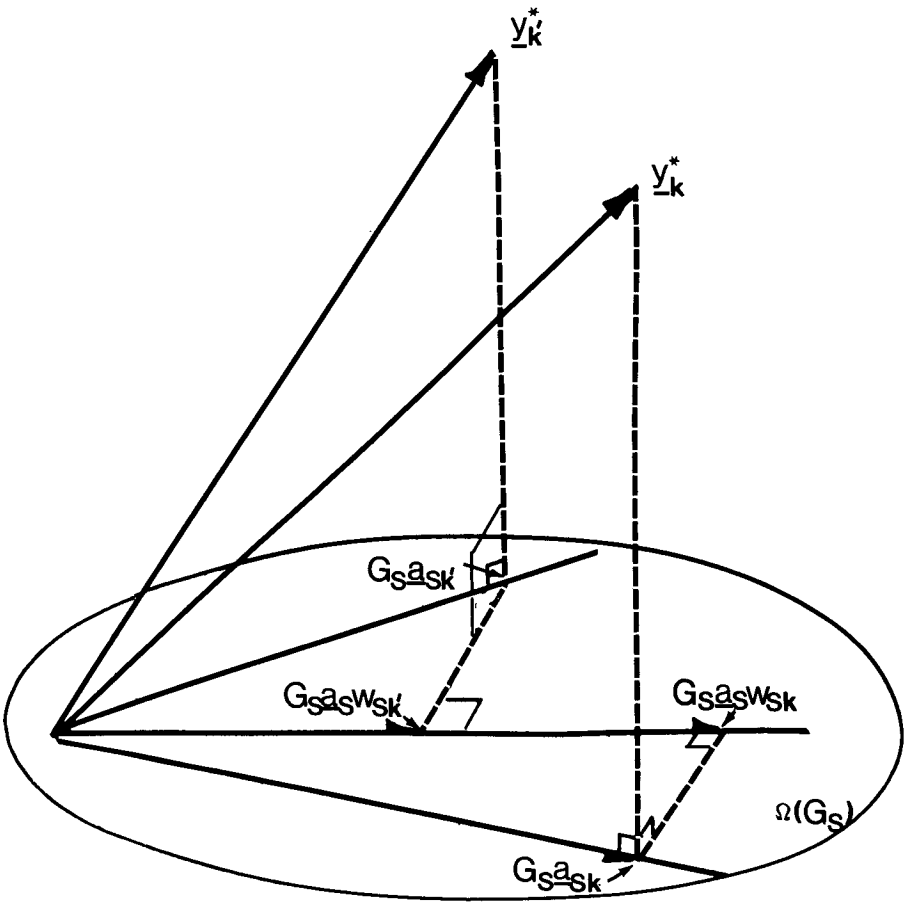

FIGURE 2.
The relationship between the weighted additive model and the simple additive model.

$$\begin{aligned} SS(\mathbf{y}_k^*) &= SS(\mathbf{G}_s\mathbf{a}_{sk}) + SS(\mathbf{y}_k^* - \mathbf{G}_s\mathbf{a}_{sk}) \\ &= SS(\mathbf{G}_s\mathbf{a}_s w_{sk}) + SS(\mathbf{G}_s\mathbf{a}_{sk} - \mathbf{G}_s\mathbf{a}_s w_{sk}) + SS(\mathbf{y}_k^* - \mathbf{G}_s\mathbf{a}_{sk}) \\ &= SS(\mathbf{G}_s\mathbf{a}_s w_{sk}) + SS(\mathbf{y}_k^* - \mathbf{G}_s\mathbf{a}_s w_{sk}), \end{aligned}$$

where $SS(\mathbf{v})$ indicates the sum of squares due to vector $\mathbf{v}$ (i.e., $SS(\mathbf{v}) = \mathbf{v}'\mathbf{v}$). The above equation shows that the total sum of squares due to $\mathbf{y}_k^*$ can be decomposed in at least three different ways: the SS due to the simple additive model, $SS(\mathbf{G}_s\mathbf{a}_{sk})$, plus its residual, $SS(\mathbf{y}_k^* - \mathbf{G}_s\mathbf{a}_{sk})$; the SS due to the weighted additive model, $SS(\mathbf{G}_s\mathbf{a}_s w_{sk})$, plus its residual, $SS(\mathbf{y}_k^* - \mathbf{G}_s\mathbf{a}_s w_{sk})$; and $SS(\mathbf{G}_s\mathbf{a}_s w_{sk})$ plus $SS(\mathbf{y}_k^* - \mathbf{G}_s\mathbf{a}_{sk})$ plus $SS(\mathbf{G}_s\mathbf{a}_{sk} - \mathbf{G}_s\mathbf{a}_s w_{sk})$. This last SS represents the sum of squares which can be accounted for by the simple additive model but not by the weighted additive model. That is, it represents the differential predictability between the two models.

In the special case of $N = 1$, the weighted additive model and the simple additive model are equivalent in every respect.

3. *Algorithm*

Optimization Criterion

Given a set of observations $\mathbf{o}_k$ ($k = 1, \cdots, N$), and design matrices $\mathbf{G}_s$ ($s \in S$), we estimate additive effects $\mathbf{a}_s$ ($s \in S$), and individual differences weights w_{sk} ($s \in S$; $k = 1, \cdots, N$)

subject to some optimality consideration. The optimally scaled data $\mathbf{y}_k^*$ ($k = 1, \cdots, N$) must also be obtained under the same optimality consideration. The optimization criterion we choose to minimize is

$$Q^{(R)} = \sum_{k=1}^{N} Q_k^{(R)}, \tag{10}$$

where

$$Q_k^{(R)} = (\hat{\mathbf{y}}_k - \mathbf{y}_k^*)'(\hat{\mathbf{y}}_k - \mathbf{y}_k^*). \tag{11}$$

However, $Q^{(R)}$ can be made identically zero by setting $\hat{\mathbf{y}}_k = \mathbf{y}_k^* = \mathbf{0}_n$ (the n-component zero vector) when the data are nonmetric (i.e., either nominal or ordinal). In order to avoid this meaningless solution, $Q^{(R)}$ should be minimized under an appropriate normalization restriction. The conventional optimization scheme (e.g., the method of steepest descent) typically requires the direct optimization of a normalized loss function corresponding to (10). With an ALS procedure, however, it has been shown [de Leeuw, et al, 1976; de Leeuw, Note 1] that a normalization restriction may be incorporated in the procedure by actually performing the normalization on whatever arguments are involved in the normalization restriction (optimally scaled data in the present case) at any step of each iteration. Otherwise the optimization procedure can proceed just as if we were optimizing the unnormalized loss function defined in (10). This property of ALS procedures is very convenient since we do not have to work on the normalized loss function directly, and since the partial derivatives of the normalized loss function are considerably more complicated than those of the unnormalized function. (They are not easily solvable in an explicit form, even for a subset of the parameters.) To be precise, we can optimize

$$Q^{(N)} = \sum_{k=1}^{N} \left(\frac{Q_k^{(R)}}{Q_k^{(D)}} \right) \tag{12}$$

where $Q_k^{(D)} = \mathbf{y}_k^{*\prime}\mathbf{y}_k^*$ by normalizing $\mathbf{y}_k^*$ so that $Q_k^{(D)} = 1$ for each k. It can be readily seen that, under this normalization convention, $Q^{(R)} = Q^{(N)}$. As noted earlier, in the weighted additive model it is assumed that the model predictions and optimally scaled data constitute individual-wise interval (affine-invariant) scales. Hence the data are comparable only within each individual. That is, the data are considered matrix conditional in Takane, Young and de Leeuw's [1977] terminology. The normalization within each individual is necessary to avoid a type of degeneracy in solutions discussed by Roskam [1968]. Since the optimally scaled data are centered for each individual, the minimization of $Q^{(N)}$ is equivalent to that of

$$Q^{(C)} = \sum_{k=1}^{N} \left(\frac{Q_k^{(R)}}{Q_k^{(D')}} \right), \tag{13}$$

where $Q_k^{(D')} = (\mathbf{y}_k^* - \bar{\mathbf{y}}_k^*)'(\mathbf{y}_k^* - \bar{\mathbf{y}}_k^*)$, and $\bar{\mathbf{y}}_k^* = 1/n(\mathbf{1}_n'\mathbf{y}_k^*\mathbf{1}_n)$ (which is a zero vector in the present case).

General Algorithmic Flow

The general framework of the algorithm construction is based on the alternating least squares principle, in which the unknown model parameters are partitioned into several subsets, and are estimated separately for each subset while temporarily assuming that the parameters in the other subsets have known values. The crucial point is that the estimates in each subset must be least squares estimates derived from a single common least squares optimization criterion. The estimation procedure is iterated until convergence is reached.

The WADDALS procedure consists of two major phases and two minor phases. The

two major phases are the model estimation phase, which in turn consists of two subphases, the estimation of additive effects and the estimation of individual differences weights, and the optimal scaling phase. The two minor phases are the initialization phase and the termination phase. The algorithm is similar to ADDALS for the simple additive model except that in the present case we have to estimate individual differences weights as well as effects of additive factors.

We first illustrate the general algorithmic flow of WADDALS, and then present a slightly more involved account of each step in the next section.

(Step 1): Initialize optimally scaled data.
(Step 2): Obtain initial estimates of additive effects.
(Step 3): Estimate individual differences weights.
(Step 4): (Re−) estimate additive effects.
(Step 5): Calculate model predictions and obtain optimally scaled data.
(Step 6): Check convergence. If converged, stop, or else go to (Step 3).

Details of the Algorithm

For simplicity we describe the algorithm for the two-factor design. The algorithm can be readily extended to higher-factor designs.

Let

$$\mathbf{y}^{*\prime} = (\mathbf{y}_1^{*\prime}, \mathbf{y}_2^{*\prime}, \cdots, \mathbf{y}_N^{*\prime}), \qquad \hat{\mathbf{y}}' = (\hat{\mathbf{y}}_1', \hat{\mathbf{y}}_2', \cdots, \hat{\mathbf{y}}_N'),$$

and

$$\mathbf{w}_s' = (w_{s1}, w_{s2}, \cdots, w_{sN}), \qquad s \in S = \{A, B\}.$$

Then the weighted additive model can be stated as

$$\mathbf{y}^* \cong \hat{\mathbf{y}} = (\mathbf{w}_A \otimes \mathbf{G}_A)\mathbf{a}_A + (\mathbf{w}_B \otimes \mathbf{G}_B)\mathbf{a}_B, \tag{6'}$$

for the two-factor design where $\otimes$ indicates a Kronecker product [see, for example, Graybill, 1969, or Rao, 1973]. This form of the statement of the model will be useful in the following discussion.

(Step 1): As in all of our previous ALS procedures, we set $\mathbf{y}_k^* = \mathbf{o}_k$ $(k = 1, \cdots, N)$ for initialization. For categories of nominal variables, for which initial values are not externally provided, we assign arbitrary numbers. These initial (optimally scaled) data are then centered and normalized within individuals.

(Step 2): Initial estimates of additive effects may be obtained by temporarily regarding $\mathbf{w}_s = \mathbf{1}_N$ for $s \in S$.

Define

$$\mathbf{G}^* = \mathbf{1}_N \otimes \mathbf{G}, \quad \text{where}$$

$$\mathbf{G} = [\mathbf{G}_A, \mathbf{G}_B], \quad \text{and}$$

$$\mathbf{a}' = (\mathbf{a}_A', \mathbf{a}_B')$$

A least squares estimate of $\mathbf{a}$ [in the sense of (10)] is given by

$$\hat{\mathbf{a}} = (\mathbf{G}^{*\prime}\mathbf{G}^*)^-\mathbf{G}^{*\prime}\mathbf{y}^*,$$

where $(\mathbf{G}^{*\prime}\mathbf{G}^*)^-$ is a generalized inverse of $\mathbf{G}^{*\prime}\mathbf{G}^*$. In order to identify a unique solution, it is convenient to impose the restriction that

$$\mathbf{a}_s'\mathbf{G}_s^{*\prime}\mathbf{1}_{N^*} = N\,\mathbf{a}_s'\,\mathbf{G}_s'\mathbf{1}_N = \frac{Nn}{n_s}\,\mathbf{a}_s'1_{n_s} = 0 \tag{14}$$

for $s \in S$, where $\mathbf{G}_s^* = \mathbf{1}_N \otimes \mathbf{G}_s$ ($s \in S$) and $\mathbf{1}_{N^*} = \mathbf{1}_N \otimes \mathbf{1}_n$ (i.e., $N^* = N \times n$). We then normalize $\hat{\mathbf{a}}_s$ so that it satisfies the normalization restriction [Eq. (7)], and adjust the length of the weight vector. The elements of $\hat{\mathbf{a}}_s$ may be subjected to an order restriction, as given in de Leeuw et al. [1976]. Note that under restriction (14) the least squares estimate of a can be obtained by simply taking appropriate means of the elements of $\mathbf{y}^*$ (i.e.,

$$\hat{\mathbf{a}}_s = \frac{1}{N}(\mathbf{G}_s'\mathbf{G}_s)^{-1}(\mathbf{G}_s' \otimes \mathbf{1}_N')\mathbf{y}^*$$

with $\mathbf{G}_s'\mathbf{G}_s$ being diagonal).

(Step 3): Let

$$\mathbf{Y}^* = [\mathbf{y}_1^*, \mathbf{y}_2^*, \cdots, \mathbf{y}_N^*],$$

$$\mathbf{X} = [\mathbf{x}_A, \mathbf{x}_B], \quad \text{where}$$

$$\mathbf{x}_s = \mathbf{G}_s\mathbf{a}_s \quad \text{for} \quad s \in S, \quad \text{and}$$

$$\mathbf{W}' = [\mathbf{w}_A, \mathbf{w}_B].$$

The estimate of $\mathbf{W}$, for fixed $\mathbf{Y}^*$ and $\mathbf{a}_s$ ($s \in S$), is given by

$$\hat{\mathbf{W}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}^*. \tag{15}$$

[Note that $Q^{(R)} = \text{tr}(\mathbf{Y}^* - \mathbf{X}\mathbf{W})'(\mathbf{Y}^* - \mathbf{X}\mathbf{W})$.] Nonnegativity restrictions may be imposed on the weight estimates. The treatment of the restrictions within the ALS framework will be discussed in the next section. Again the inversion of $\mathbf{X}'\mathbf{X}$ is trivial in (15), since it is diagonal.

(Step 4): Define

$$\mathbf{G}_w^* = [\mathbf{G}_{Aw}^*, \mathbf{G}_{Bw}^*],$$

where $\mathbf{G}_{sw}^* = \mathbf{w}_s \otimes \mathbf{G}_s$ for $s \in S$. The least squares estimate of $\mathbf{a}$ for fixed $\mathbf{w}_s$ ($s \in S$) and $\mathbf{y}^*$ is given by

$$\hat{\mathbf{a}} = (\mathbf{G}_w^{*\prime}\mathbf{G}_w^*)^{-1}\mathbf{G}_w^{*\prime}\mathbf{y}^*. \tag{16}$$

Note that the regular inverse can be defined for $\mathbf{G}_w^{*\prime}\mathbf{G}_w^*$, and that $\hat{\mathbf{a}}$ is uniquely determined without imposing any additional restriction [such as (14)] unless $\mathbf{w}_s$ is a constant vector for $s \in S$. We normalize $\hat{\mathbf{a}}_s$ and adjust the length of $\mathbf{w}_s$ accordingly. Again, order restrictions may be imposed on the elements of $\mathbf{a}_s$. The way in which order restrictions are incorporated will be discussed in the next section. Under the separability conditions (see the next section) (16) reduces to (18), and since $\mathbf{G}_{sw}^{*\prime}\mathbf{G}_{sw}^*$ ($s \in S$) is diagonal, $\hat{\mathbf{a}}$ in the above equation can be obtained without explicit matrix inversions.

(Step 5): Model predictions are calculated through (6) using the current estimates of the model parameters. We now wish to calculate the optimally scaled data (the least squares estimates of the model predictions under the measurement restrictions which are implied by the measurement characteristics of the observations). The specific optimal scaling procedure depends on the assumed measurement level of the observations and the assumed (discrete or continuous) measurement process. For example, if observations are ordinal we apply Kruskal's least squares monotonic transformation [Kruskal, 1964] with either primary (in case of continuous process) or secondary (in case of discrete process) approach to ties. Or if observed data are nominal, we apply the least squares nominal transformation [de Leeuw et al, 1976]. Those concepts, the mathematical operations and their justification, have been fully discussed elsewhere [see, in particular, Young, de Leeuw and Takane, 1979], and therefore will not be repeated here.

Since observations are assumed comparable only within individuals, the optimal

scaling is performed separately for each individual. Once the optimally scaled data are obtained, they are then normalized. (If $\mathbf{y}_k^*$ is initially centered, it is kept centered throughout all iterations.)

(Step 6): We evaluate the goodness of fit by (10). If the improvement of fit from the previous iteration is smaller than a prescribed value, the procedure is terminated. Otherwise, the next iteration is commenced. This termination criterion is justified by the monotone convergence property of ALS procedures [Zangwill, 1969; de Leeuw, Young & Takane, 1976].

Principle of Separate Optimization

In this section we deal with a special problem related to the treatment of external constraints on parameters within the ALS framework. Our exposition is terse. However, the nonmathematical reader may skip the entire section without loss of continuity.

External constraints on parameters, such as the nonnegativity restriction on w_{sk} or the order restriction on $\mathbf{a}_s$, may sometimes be quite easily incorporated in the ALS optimization framework. The ease with which they can be incorporated, however, critically depends on the notion of the "separability" of parameters to be defined below.

Definition 1: Separability of parameters with respect to a specific optimization criterion. Partitioned sets of parameters are said to be separable with respect to a specific optimization criterion when a monotonically increasing function thereof can be decomposed into additive components, each of which is a function of only a subset of the parameters.

The optimal set of parameter estimates in this case can be obtained by optimizing each of the additive components separately with respect to its arguments (*Principle of Separate Optimization*).

Definition 2: Separability of constraints. Constraint sets are said to be separable if they are expressible as a set of separate statements each of which involves only a subset of parameters. (We sometimes say that subsets of parameters are separable with respect to the set of constraints.)

If parameters are separable in both senses, (Definitions 1 and 2) and the separable subsets of parameters coincide, then the total set of the constrained parameter estimates may be obtained by separately obtaining the constrained estimates, assuming that they optimize each separable component of the global optimization criterion under the corresponding constraints.

Proposition 1: The sets of $\{\mathbf{w}_{Ak}, \mathbf{w}_{Bk}\}$, or more generally $\{\mathbf{w}_{sk}, s \in S\}$, $k = 1, \cdots, N$ are separable with respect to the loss function (10).

Assumption 1: The design in the model (6) is balanced.

That is,

$$\mathbf{G}_A'\mathbf{G}_B = \frac{1}{n}\mathbf{G}_A'\mathbf{1}_n\mathbf{1}_n'\mathbf{G}_B \tag{17}$$

holds for all A, $B \in S$ $(A \neq B)$.

Assumption 2: The restriction (14) holds for $s \in S$.

Proposition 2: The sets of $\{w_{sk}\}$, $s \in S$ for any k are separable under Assumptions 1 and 2. Note that $X'X = D$ (diagonal) for X defined in Step 3.

Propositions 1 and 2 establish the separability of each w_{sk} under Assumptions 1 and 2. Since the nonnegativity restriction is imposed separately on w_{sk}, the constrained estimate can be obtained separately for each w_{sk} by minimizing each component of the loss function (10) related to w_{sk}. Furthermore, it can be shown that each separable component of the loss function is a quadratic function of w_{sk}.

Proposition 3: Suppose $Q(\alpha)$ is a positive quadratic function in a single parameter α. The minimum of Q, under the restriction that $\alpha^- \leq \alpha \leq \alpha^+$, is attained at

(i) $\alpha = \alpha^-$ if $\hat{\alpha} < \alpha^-$, (Case *A*)

(ii) $\alpha = \hat{\alpha}$ if $\alpha^- < \hat{\alpha} < \alpha^+$, (Case *B*)

(iii) $\alpha = \alpha^+$ if $\hat{\alpha} > \alpha^+$, (Case *C*)

where $\hat{\alpha}$ gives the unconstrained minimum of Q (See Figure 3).

Thus, the constrained estimate of w_{sk} is given by $w_{sk} = \hat{w}_{sk}$ if $\hat{w}_{sk} \geq 0$ where $\hat{w}_{sk}$ is the unconstrained estimate and by $w_{sk} = 0$ if $\hat{w}_{sk} < 0$.

Proposition 4: Under Assumptions 1 and 2 additive effects in the simple additive model are separable with respect to the least squares criterion analogous to (10).

Proposition 5: Under the same assumptions additive effects in the weighted additive model are separable with respect to (10). (Proof): We would like to prove $Q^{(R)} = \sum_{s \in S} Q(\mathbf{a}_s) + c$ for some constant c, and where $Q(\mathbf{a}_s) = (\mathbf{y}^* - \mathbf{G}^*_{sw}\mathbf{a}_s)'(\mathbf{y}^* - \mathbf{G}^*_{sw}\mathbf{a}_s)$ for $s \in S$. We first establish $\mathbf{a}'_A \mathbf{G}^{*\prime}_{Aw} \mathbf{G}^*_{Bw} \mathbf{a}_B = 0$ for $A \neq B$. (The desired results immediately follow from this.) We have

$$\begin{aligned} \mathbf{G}^{*\prime}_{Aw}\mathbf{G}^*_{Bw} &= (\mathbf{w}_A \otimes \mathbf{G}_A)'(\mathbf{w}_B \otimes \mathbf{G}_B) \\ &= (\mathbf{w}'_A\mathbf{w}_B)\mathbf{G}'_A\mathbf{G}_B \\ &= \frac{1}{n}(\mathbf{w}'_A\mathbf{w}_B)\mathbf{G}'_A\mathbf{1}_n\mathbf{1}'_n\mathbf{G}_B \qquad \text{(by Assumption 1)}, \end{aligned}$$

and $\mathbf{a}'_A \mathbf{G}^{*\prime}_{Aw} \mathbf{G}^*_{Bw} \mathbf{a}_B = 0$ by noting Assumption 2. Thus, (16) is equivalent to

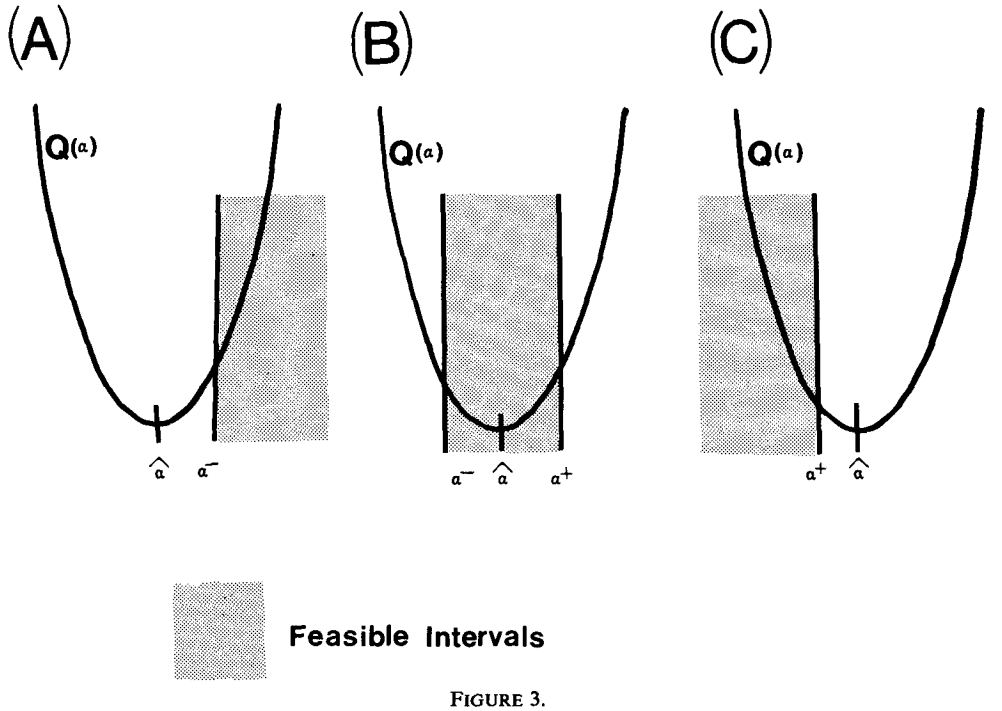

FIGURE 3.
The constrained minima of a quadratic function in a single parameter.

$$\hat{\mathbf{a}}_s = (\mathbf{G}_{sw}^{*\prime}\mathbf{G}_{sw}^{*})^{-1}\,\mathbf{G}_{sw}^{*\prime}\mathbf{y}^{*}, \qquad (s \in S), \tag{18}$$

concluding the proof. Note that the $\hat{\mathbf{a}}_s$ defined above satisfies the condition implied by Assumption 2. This can be easily shown by

$$\begin{aligned}\mathbf{a}_s'\mathbf{1}_{n_s} &= \mathbf{y}^{*\prime}\mathbf{G}_{sw}^{*\prime}(\mathbf{G}_{sw}^{*\prime}\mathbf{G}_{sw}^{*})^{-1}\mathbf{1}_{n_s} \\ &= c\mathbf{y}^{*\prime}\mathbf{G}_{sw}^{*\prime}\mathbf{1}_{n_s} \\ &= c\sum_{k=1}^{N} w_{sk}\mathbf{y}_k^{*\prime}\mathbf{1}_n = 0,\end{aligned}$$

since $\mathbf{y}_k^{*\prime}\mathbf{1}_n = 0$ for $k = 1, \cdots, N$.

We now know that each $\mathbf{a}_s$ is separable. Hence, if the monotonicity restriction is imposed separately on each $\mathbf{a}_s$, the constrained estimate (denoted by $\mathbf{a}_s^{*}$) of $\mathbf{a}_s$ which minimizes $Q^{(R)}$ is given by the constrained estimate which minimizes $Q(\mathbf{a}_s)$ for each $s \in S$. It is well-known that the constrained estimate $\mathbf{a}_s^{*}$ which minimizes $Q(\mathbf{a}_s)$ under the monotonicity restriction, is obtained by applying Kruskal's [1964] least squares monotonic transformation to the elements of $\hat{\mathbf{a}}_s$. Note that the $\mathbf{a}_s^{*}$ obtained this way also satisfies the relation implied by Assumption 2. This can be shown by noting that $\mathbf{a}_s^{*}$ can be generally written as $\mathbf{a}_s^{*} = \hat{\Pi}_{n_s}\hat{\mathbf{a}}_s$ where $\hat{\Pi}_{n_s}$ is an orthogonal projection operator with the property that $\hat{\Pi}_{n_s}\mathbf{1}_{n_s} = \mathbf{1}_{n_s}$. Thus, $\mathbf{a}_s^{*\prime}\mathbf{1}_{n_s} = \hat{\mathbf{a}}_s'\hat{\Pi}_{n_s}\mathbf{1}_{n_s} = \hat{\mathbf{a}}_s'\hat{\Pi}_{n_s} = \hat{\mathbf{a}}_s'\mathbf{1}_{n_s} = 0$. That the estimates of $\mathbf{a}_s$ ($\hat{\mathbf{a}}_s$ or $\mathbf{a}_s^{*}$) satisfy (14) is very crucial, since this is the necessary condition for the separability.

Note that the elements of $\mathbf{a}_s$ in each s are also separable, since $\mathbf{G}_{sw}^{*\prime}\mathbf{G}_{sw}^{*}$ = diagonal. However, external constraints on the element of $\mathbf{a}_s$, say a nonnegativity restriction of a_{s_i}, cannot be handled in the same way as the nonnegativity restriction on individual differences weight is handled. Setting $a_{s_i}^{*} = 0$ for which $\hat{a}_{s_i} < 0$ would destroy the relation (14) on $\mathbf{a}_s^{*}$. [Note that (14) is a nonseparable constraint on the element of $\mathbf{a}_s$.]

Proposition 6: The optimally scaled data $\mathbf{y}_k^{*}$ ($k = 1), \cdots, N$) are separable under any circumstance.

This last proposition is rather obvious from the definition of (10). Its importance, however, should not be overlooked since it provides a basic rationale for applying the optimal data transformation method separately to each individual's data.

The notion of separability plays an important role in the construction of the current algorithm. Thus, experiments should be designed in such a way that the conditions for separability may be satisfied as much as possible. The only serious problem in this regard seems to be Assumption 1 (the assumption that the design is balanced). As we have seen, Propositions 2, 3 and 4 (the separability of w_{sk} for $s \in S$ and the separability of additive effects in the simple and the weighted additive models) are critically dependent on this assumption.

There are basically two ways to deal with the problem of unbalanced designs. One is simply to give up all of the convenient properties which follow from separability. The ALS procedure is flexible enough to do the estimation without relying on the separability properties; we can use the MORALS [Young et al, 1976] type of sequential estimation procedure, although the estimation procedure itself as well as the incorporation of external constraints would no longer be as simple as before. Lawson and Hanson [1974] also discuss the general problem of least squares under arbitrary linear equality and inequality constraints.

The other way to deal with unbalanced designs, which is more in tune with the above discussion on separability, is to force the design to be balanced by treating missing cells (cells which are not included in the design) as if they are missing observations (i.e., cells were actually included in the design, but no observations were made for the cells for some

incidental reasons). In this case we simply set optimally scaled data corresponding to missing observations equal to their model predictions. In other words we regard missing observations as having no constraining power on the solution. Since the design is considered balanced, the separability properties hold just as when there are no missing observations.

Note that in either case the stress (the square root of $Q^{(N)}$ defined in (12) at an optimal point) would tend to show a spuriously improving fit as the number of missing observations increases. This, though it may seem strange at first glance, is quite natural when we realize that missing observations convey no information about the model, and consequently can always be perfectly fitted by the model (or for that matter by *any* model). It is also consistent with the behavior of stress in general that it tends to take a smaller value for a smaller design (i.e., when the effective number of observations/the effective number of parameters to be estimated is smaller). Of course, this behavior of stress is an undesirable property of a goodness of fit index. We have not investigated the effect of missing data on stress systematically. The WADDALS program, however, has been written in such a way that Monte Carlo experiments can be readily run to investigate this problem by just specifying the assumed proportions of missing data elements. (See the discussion by Nishisato [1979] for the same problem in a somewhat different context.)

4. Evaluations

In this section we present some empirical evidence in support of the theoretical development of the WADDALS procedure outlined in the preceding sections. We first evaluate the performance of the procedure using Monte Carlo data focusing on its ability to recover "true" underlying structures and on its monotone-invariance property (when the data are assumed ordinal). We then describe a practical data analytic situation to demonstrate the use of the current procedure.

Goodness of Recovery of the Original Information

Preliminary considerations. For practical applications of nonmetric procedures it is important to set rough guidelines as to their limitations. For example, a reasonably large number of observations are necessary to obtain reasonably accurate estimates of parameters. But specifically how many observations do we need? This is the basic question we attempt to answer in this section in reference to the WADDALS procedure.

We explore the problem using a Monte Carlo technique. We first generate a set of "true" underlying structures (populations), calculate model predictions, and add random errors to obtain a set of "observed" data. The "observed" data are then submitted to the WADDALS procedure to see how well it can recover the original information. The goodness of recovery is measured in terms of some agreement measure between "true" populations (parameters) and estimated values. (The agreement measure we use depends on the scale properties of the model parameters.)

There are two distinct kinds of parameters in the weighted additive model. Whereas the individual differences weights are determined up to a ratio scale, the additive effects are only determined up to a joint interval scale. For ratio measurement we may simply take the normalized sum of cross products as an agreement measure. Let $\mathbf{w}_k$ and $\hat{\mathbf{w}}_k$ be vectors of "true" and estimated weights for individual k. Then the agreement between $\mathbf{w}_k$ and $\hat{\mathbf{w}}_k$ can be measured by

$$r_k^{(w)} = \mathbf{w}_k'\hat{\mathbf{w}}_k / ||\mathbf{w}_k|| \cdot ||\hat{\mathbf{w}}_k||$$

where $||\mathbf{w}_k|| = (\mathbf{w}_k'\mathbf{w}_k)^{1/2}$. The $r_k^{(w)}$ defined above corresponds to the cosine of the angle between the two vectors. The smaller the discrepancy between two vectors, the larger the

value of $r_k^{(w)}$. We may take a root mean square of $r_k^{(w)}$ to define the overall measure of agreement across individuals; i.e.,

$$r^{(w)} = \left\{ \frac{\sum [r_k^{(w)}]^2}{N} \right\}^{1/2}.$$

For interval scale measurement the usual product moment correlation coefficient serves as a natural measure of agreement between two sets of numbers. Again we may use the root mean square correlation to indicate overall agreement. This overall agreement between "true" and estimated additive effects is denoted by $r^{(a)}$. Since there is no simple relationship between $r^{(w)}$ and $r^{(a)}$, the goodness of recovery will be treated separately for the individual differences weights and for the effects of additive factors.

We investigate the goodness of recovery of the original information as a function of both the number of observations and the magnitude of random errors. The number of observations in the present case, however, is directly related to such design constraints as the number of factors, the number of levels in each factor and the number of individuals. In this study we limit our attention to the two-factor case in which the number of levels in each factor is varied in three steps: 4, 6 and 10. The number of individuals is also varied in three steps: 1, 4 and 10. Then, $n_A + n_B + 2N$ uniform random numbers are generated for each data set (n_A for $\mathbf{a}_A$, n_B for $\mathbf{a}_B$ and $2N$ for w_{sk}, where $s = A, B$ and $k = 1, \cdots, N$). These "true" parameter values are then subjected to suitable normalizations. In particular, the individual differences weights are required to be $||\mathbf{w}_k|| = 1$ for each individual. Model predictions are calculated according to the model. Then independent normal errors are added to the predictions to obtain the "observed" data. Three levels of random error (σ_e) are considered: 0.25, 0.5 and 1.0. We thus have a $3 \times 3 \times 3 \times 3$ (3 levels of n_A, 3 levels of n_B, 3 levels of N and 3 levels of σ_e) design for this study. Finally, ten independent replications are obtained under each condition, giving a total of 810 data sets to be analyzed by WADDALS.

Individual differences weights. Only four of the 810 $r^{(w)}$ were lower than .900. All four occurred in the single individual cases ($N = 1$). A vast majority of $r^{(w)}$ were higher than .950. We do not present a complete table of mean stress values here. Instead, a regression equation has been developed which can predict the expected $r^{(w)}$ given n_A, n_B, N and σ_e [Spence, 1979]:

$$(19) \qquad r^{(w)} = .947 - .218 \times 10^{-1} \times \sigma_e + .656 \times 10^{-1} \times n_{A+B} - .177 \times 10^{-3} \times n^2_{A+B},$$

where $n_{A+B} = n_A + n_B$. (The standard error for this regression was .017.) Note that N, the number of individuals, has little effect on $r^{(w)}$, and consequently is not included in the above formula, although the smaller N, as indicated by several instances of exceptionally low $r^{(w)}$ ($< .900$), increases the probability of degenerate solutions. This is because larger N does not necessarily increase the number of observations/number of parameters ratio for the weight estimates.

Additive effects. There were 12 cases in which $r^{(a)}$ was lower than .700, 11 of which occurred when $N = 1$. In all of the 12 cases the cause for the exceptionally low $r^{(a)}$ was rather trivial; the generated individual differences weights were disproportionately large on one factor and disproportionately small on the other. The additive effects of a factor to which disproportionately small weights were attached were very poorly recovered. Of course, as N increases the chances will be less that large weights are consistently given to one of the factors for all N individuals. In fact, for $N = 10$, $r^{(a)}$ were all above .900. The prediction formula for $r^{(a)}$ analogous to (19) is given by

$$(20) \qquad r^{(a)} = .986 - .155 \times \sigma_e + .125 \times 10^{-1} \times (\ln N)^{1/2} + .829 \times 10^{-1} \times \sigma_e \times (\ln N)^{1/2},$$

with the standard error of .056. As can be seen, there is an interaction between σ_e and N. Note that neither n_A nor n_B enter into the above equation.

Discussion. A general conclusion we may draw from the above analysis is that the WADDALS procedure is capable of obtaining reasonably accurate estimates ($r^{(w)} > .950$ and $r^{(a)} > .900$) of model parameters, provided that the situation is not exceedingly unfavorable. We should certainly avoid the case in which one or more factors receive near zero weights for all individuals since the estimates of additive effects for the factors with (near) zero weights are extremely unreliable. Yet the overall predictability of the model is relatively unimpaired, simply because those unreliable estimates receive near zero weights in deriving model predictions, and the additive effects are still accurately estimated for the other factors (which receive relatively large weights).

One major difficulty in the practical use of formulae (19) and (20) is that σ_e is usually not available. We will discuss how we can cope with this problem in a later section.

Monotone-Invariance Property

Data generation. One of the critical features of nonmetric procedures is that they are able to obtain invariant results over the transformations of the data which are admissible for a particular set of measurement characteristics. For example, if the procedure is ordinal, it should obtain invariant results over any monotonic transformations of the original data. The purpose of the second Monte Carlo study is to demonstrate this monotone-invariance property with the WADDALS procedure. Note that in the Monte Carlo study reported in the previous section no monotonic distortions were applied, though the data were assumed to be ordinal. Thus, the results obtained in the previous section might have been unduly favorable because of the favorable initializations. (Our initialization method is most appropriate when there is no monotonic distortion.) However, if we can show the monotone-invariance property of the WADDALS procedure, then we are indeed demonstrating the ability of WADDALS to recover the original information under the general circumstance.

For this study we hypothesize a two-factor structure with $n_A = 5$, $n_B = 6$ and $N = 10$, which are fixed throughout the experiment. From the result in the previous section we are reasonably confident that this design would lead to fairly accurate estimates of model parameters, when no monotonic distortions are applied to the data. We have chosen this case because it would be meaningless to discuss the monotone-invariance property when the original information is not appropriately recovered even in nondistorted situations.

The data without monotonic distortions are generated by

$$y_{ijk}^{(m)} = a_{A_i} w_{Ak} + a_{B_j} w_{Bk} + \sigma_e^{(m)} e_{ijk},$$

where A_i indicates the i^{th} level of factor A, B_j the j^{th} level of factor B, and $\sigma_e^{(m)}$ represents the m^{th} level of random error, and where $e_{ijk} \sim N(0, 1)$. We consider three levels of random error: ($R1$) $\sigma_e^{(1)} = 0.0$ (error-free case), ($R2$) $\sigma_e^{(2)} = 0.1$ (small error), and ($R3$) $\sigma_e^{(3)} = 0.3$ (moderate error). Parameter values ($\mathbf{a}_A$, $\mathbf{a}_B$ and $\mathbf{w}_k$) are generated by uniform random numbers. Note that the same set of parameter values and the same set of random errors (e_{ijk}) are used for all error levels so that ($R1$), ($R2$) and ($R3$) only differ with regard to the magnitude of random error ($\sigma_e^{(m)}$), and are equivalent in all other respects.

These three sets of data are then monotonically distorted in the following manner:

(S1) $${}^1o_{ijk}^{(m)} = y_{ijk}^{(m)} \quad \text{(identity)},$$

(S2) $${}^2o_{ijk}^{(m)} = y_{ijk}^{(m)2} \quad \text{(squaring)},$$

(S3) $${}^3o_{ijk}^{(m)} = (y_{ijk}^{(m)} - \bar{y}_k^{(m)})^2 \operatorname{signum}(y_{ijk}^{(m)} - \bar{y}_k^{(m)}) \quad \text{(inverse sigmoid)},$$

for $m = 1, 2$ and 3, where $\bar{y}_k^{(m)} = \Sigma_i \Sigma_j y_{ijk}^{(m)}/n_A n_B$ (the average of $y_{ijk}^{(m)}$ within individuals). The (S1) corresponds to the nondistorted case. A total of nine sets of data are generated by combining three levels of random error and three types of monotonic distortion.

Results and discussion. The nine sets of data were analyzed by WADDALS under the ordinal measurement assumption. As anticipated, the recovered optimal data transformations were very much like the original transformations. As an example, the plot of nonlinear fit (observed data on x-axis vs model predictions on y-axis) and the recovered montonic transformation (observed data on x-axis vs optimally scaled data on y-axis) are shown in Figures 4 and 5 for (R1, S3) and (R3, S3), respectively. The nonlinear fit is shown by circular dots with larger dots indicating more observation points falling on the spot, while the recovered transformation is indicated by connected line segments. It can be observed that the anticipated data transformations have been recovered very well (although they naturally deteriorate somewhat as the amount of random error increases), and that the goodness of fit gets worse as random error gets larger (as indicated by larger overall departures of dots in Figure 5 from the recovered transformation).

Table 1 summarizes the stress values (the square root of $Q^{(R)}$) at convergence (convergence is assumed to have been reached when the improvement in fit from the previous iteration is less than .0005). It can be observed that stress gets larger as the amount of random error increases (which is expected from our previous inspection of the plots of nonlinear fit). However, the fit is remarkably stable across different types of monotonic distortions, indicating that the WADDALS procedure gives monotone-invariant results.

The monotone-invariance property of the procedure has been further confirmed by

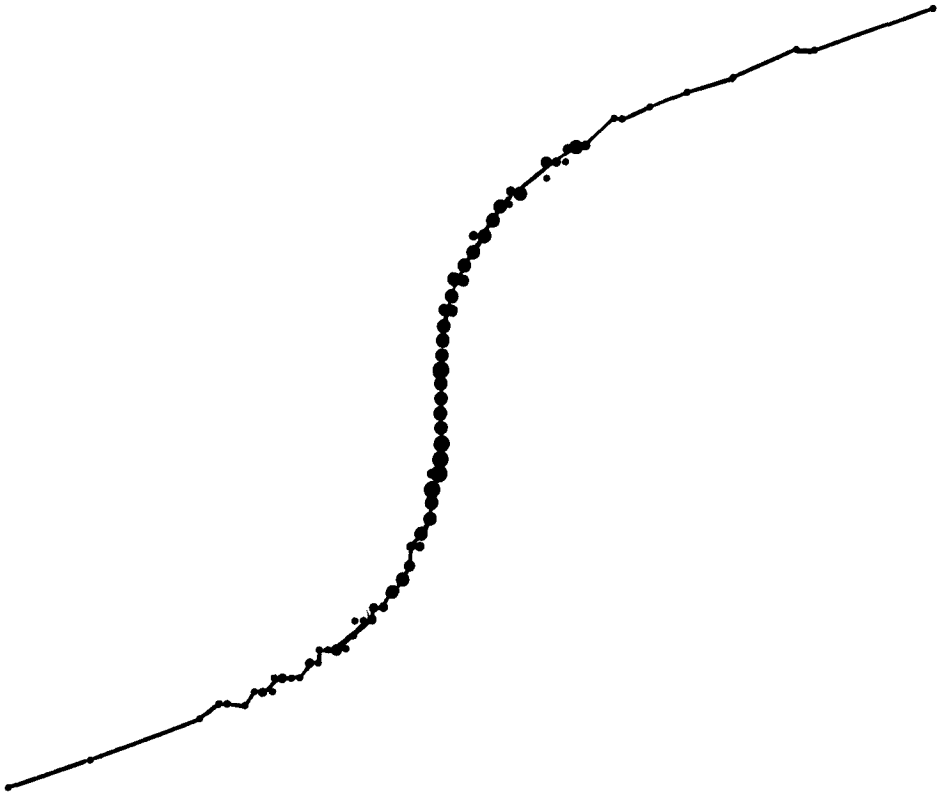

FIGURE 4.
The nonlinear fit and the recovered transformation (R1, S3).

FIGURE 5.
The nonlinear fit and the recovered transformation (R3, S3).

the fact that the estimates of model parameters were virtually indistinguishable across different types of monotonic distortions. In Figures 6, 7 and 8 we present the plots of the recovered individual differences weights and the recovered additive effects. Although the recovered parameter values deteriorate as the amount of random error increases (R1, R2 and R3; almost perfect recoveries were obtained in R1), they were nearly identical for a given level of random error. So much so that the estimates obtained under different types of monotonic distortion (S1, S2 and S3) are not explicitly distinguished in these figures.

Analysis of Kempler's Data

In this section we report an analysis of actual data. In this report a special attention will be drawn to various tests of hypotheses (model evaluations) with the WADDALS procedure. Those hypotheses include random ranking, individual-wise additivity, three-way additivity, the simple additivity and the weighted additivity vs interactions. Specifically we would like to demonstrate how Monte Carlo experiments should be set up and their results can be used for testing these hypotheses.

Data. The development of the WADDALS procedure is closely linked with an empirical observation known as the conservation of quantity in Piagetian psychology. Acquisition of the conservation of quantity may be viewed as change in children's perceptual structures. For example, younger children tend to put more emphasis on height than

TABLE 1
Stress as a function of random error levels

	Monotone distortions		
Random Errors	S1	S2	S3
R1	.000*	.009	.004
R2	.025	.026	.025
R3	.197	.197	.197

* The values reported are the square root of $Q^{(R)}$.

width when judging the largeness of rectangles, leading to a failure in conservation. They are so impressed (centrated) by the height of rectangles that they are apt to overlook other relevant aspects of stimuli. As they get older they are gradually decentrated [Liebert, Poulos & Strauss, 1974] and become able to recognize multidimensional characteristics of the stimuli.

The above contention can be paraphrased as follows. Whether the perceived area of rectangles has an additive conjoint structure (whether it is describable by the simple additive model) is a basic question in psychology. Supposing it is true, is it possible to represent developmental changes in children's perceptual structure of rectangle area judgments by the weighted additive model? Do individual differences weights attached to height decrease and those attached to width increase as the children get older? We have already discussed these problems in the introduction. We are now in the position to answer these questions on an empirical basis.

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 different age groups (1st, 3rd, 5th and 7th graders) judged each of the 100 stimuli as to whether it looked "large" or "small". (This particularly simple experimental procedure to obtain judgments was dictated by the age of the subjects.) The number of children who judged a rectangle as large was calculated for each stimulus, and used as a dependent measure indicating the largeness of the rectangle. The same experiment was repeated twice on the same sample of subjects. The aggregated data over the replicated experiments were used throughout the analysis. Whenever the reliability of the results was in question, separate analyses were performed on the two sets of data obtained in the two experiments to assess the degree of agreement between them.

Basic results. Since the weighted additive representation of the total data set presupposes the simple additive representation of each individual data set, each age group's data were first analyzed by the simple additive model. The stress values for the groups were .186, .186, .188 and .122 for the 1st, 3rd, 5th and 7th graders, respectively. The characteristic pattern in the stress values over different age groups (i.e., that the stress values are about equal for the first three age groups and a bit (but discernibly) smaller for the 7th graders) was stable across the replicated experiments. Thus, the observation that the 7th graders' responses were more consistent with the additive model than the other age groups can be considered fairly reliable (at least for the sample of subjects examined). We may reserve a complete generalization, however, since the replications were obtained from the same sample of subjects, and consequently they were not independent of each other. The joint stress value for fitting the simple additive model to the individual data set (which is equal to the root mean square of the individual stresses) was .172. The estimated additive effects were nearly linear with respect to the physical length for both height and width across all age groups. This implies that for the range of height and width levels employed in the experiments f_{Hk} and f_{Wk} in (1′) can be well approximated by $w_{Hk}f_H$ and $w_{Wk}f_W$ in (2)

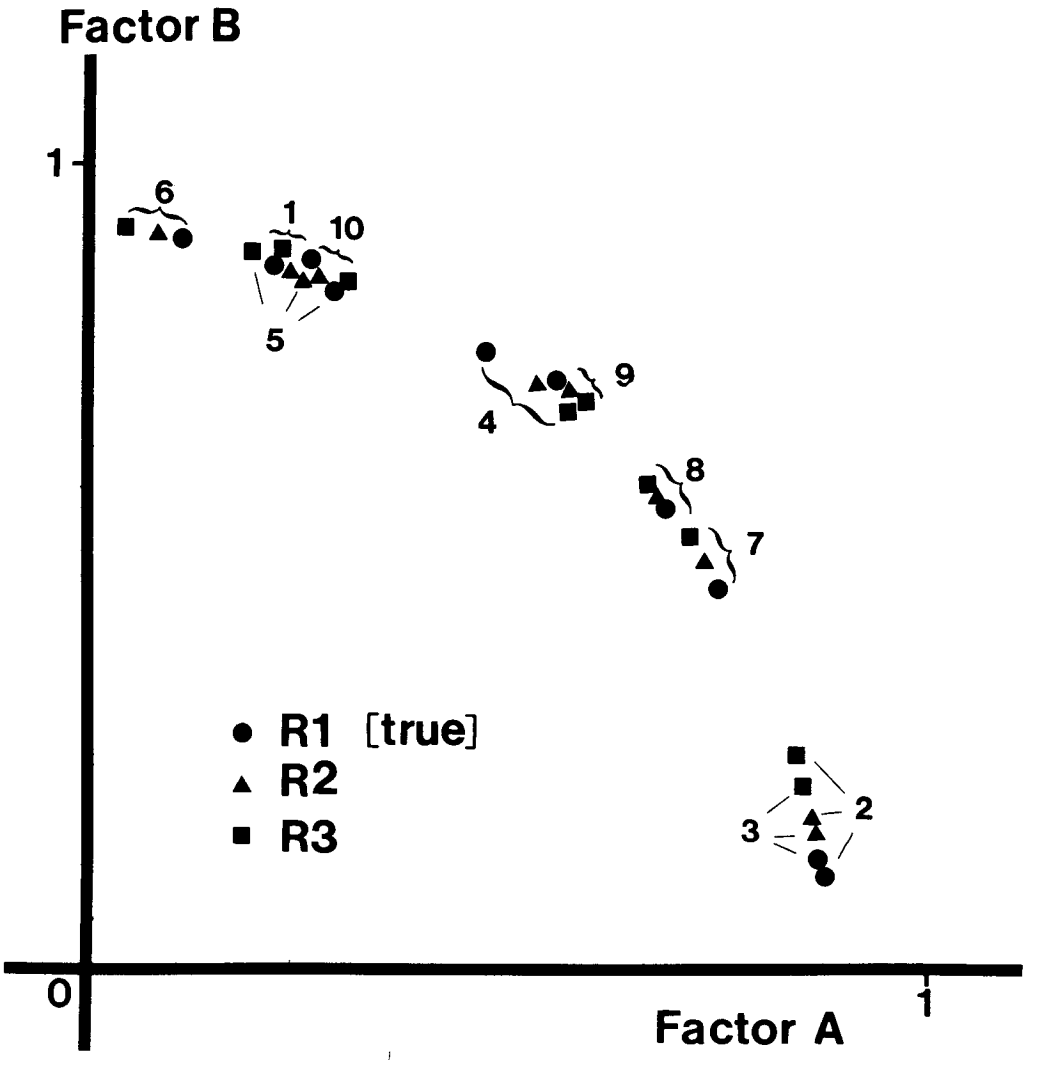

FIGURE 6.
Estimated individual differences as a function of random error levels.

respectively, for all age groups with f_H and f_W being approximately linear. Thus, there seems to be enough justification to perform the weighted additivity analysis of the total data set.

Figure 9 illustrates the estimated individual differences weights as a function of age groups. As hypothesized, the weights attached to the height of rectangles consistently decrease with age (except 5th graders), and the reverse is true for the weights attached to the width. [We note, however, that, since Kempler's data are group data (individual differences are group differences), the gradual change in the weights does not necessarily imply that the transition is gradual for each individual. The transition from the height-dominant state to the even state may be sudden for each individual, even though the proportions may change only gradually. In order to isolate the two possible interpretations, we need to obtain longitudinal data.] The estimates of additive effects were both found to be nearly linear with respect to the physical length (as expected from the above observation). The optimal transformation of the dependent variable obtained under the ordinal assumption

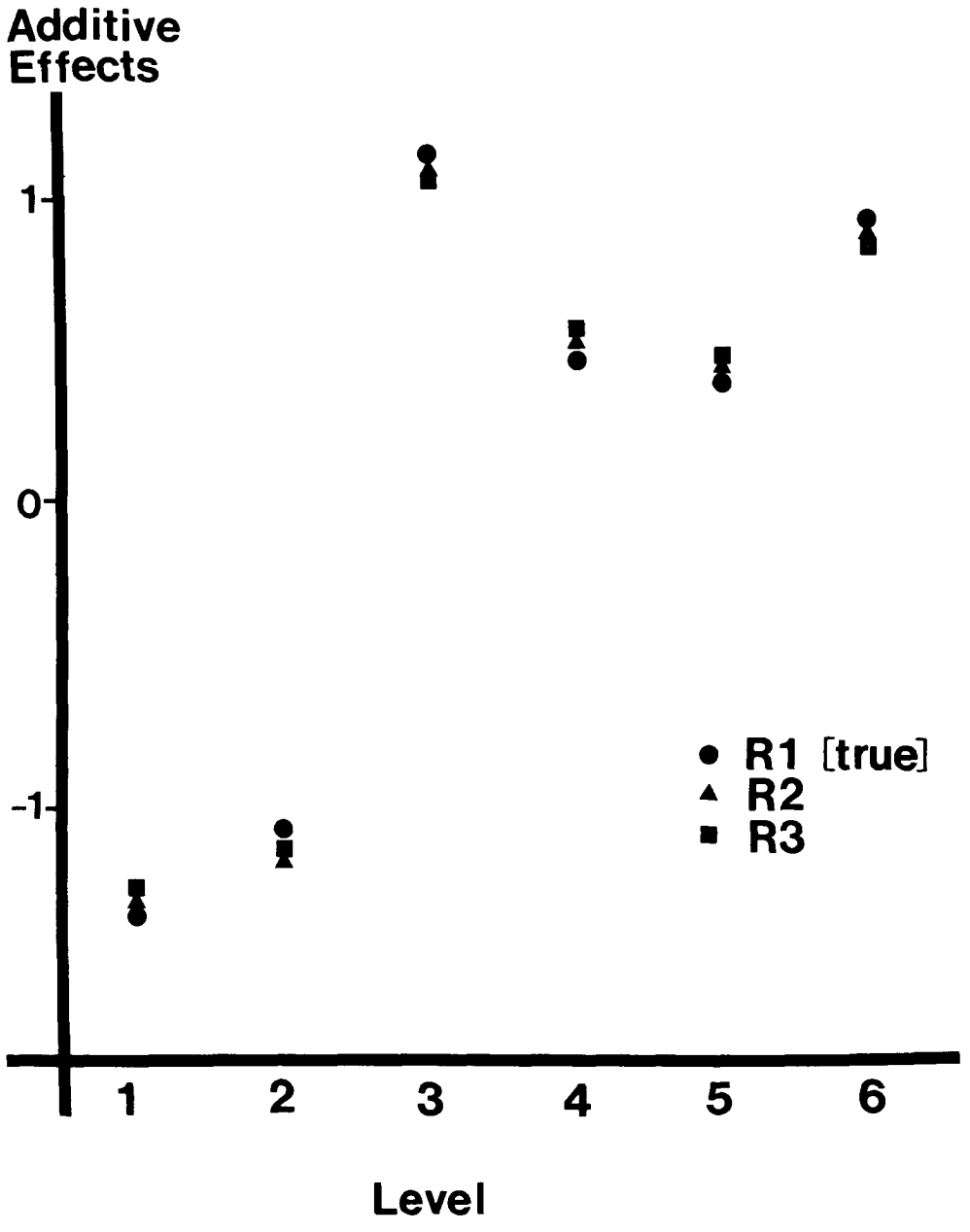

FIGURE 7.
Estimated additive effects of Factor A as a function of random error levels.

looked very much like an inverse sigmoid curve (with data on x-axis vs optimally scaled data on y-axis), a transformation typically associated with the influence of floor and ceiling effects.

In order to assess the reliability of the above results, two sets of data obtained from the two experiments were analyzed separately. The estimates of model parameters were remarkably stable over the two replications. The agreements were .99992 for $r^{(w)}$ (almost perfect agreement) and .99409 for $r^{(a)}$.

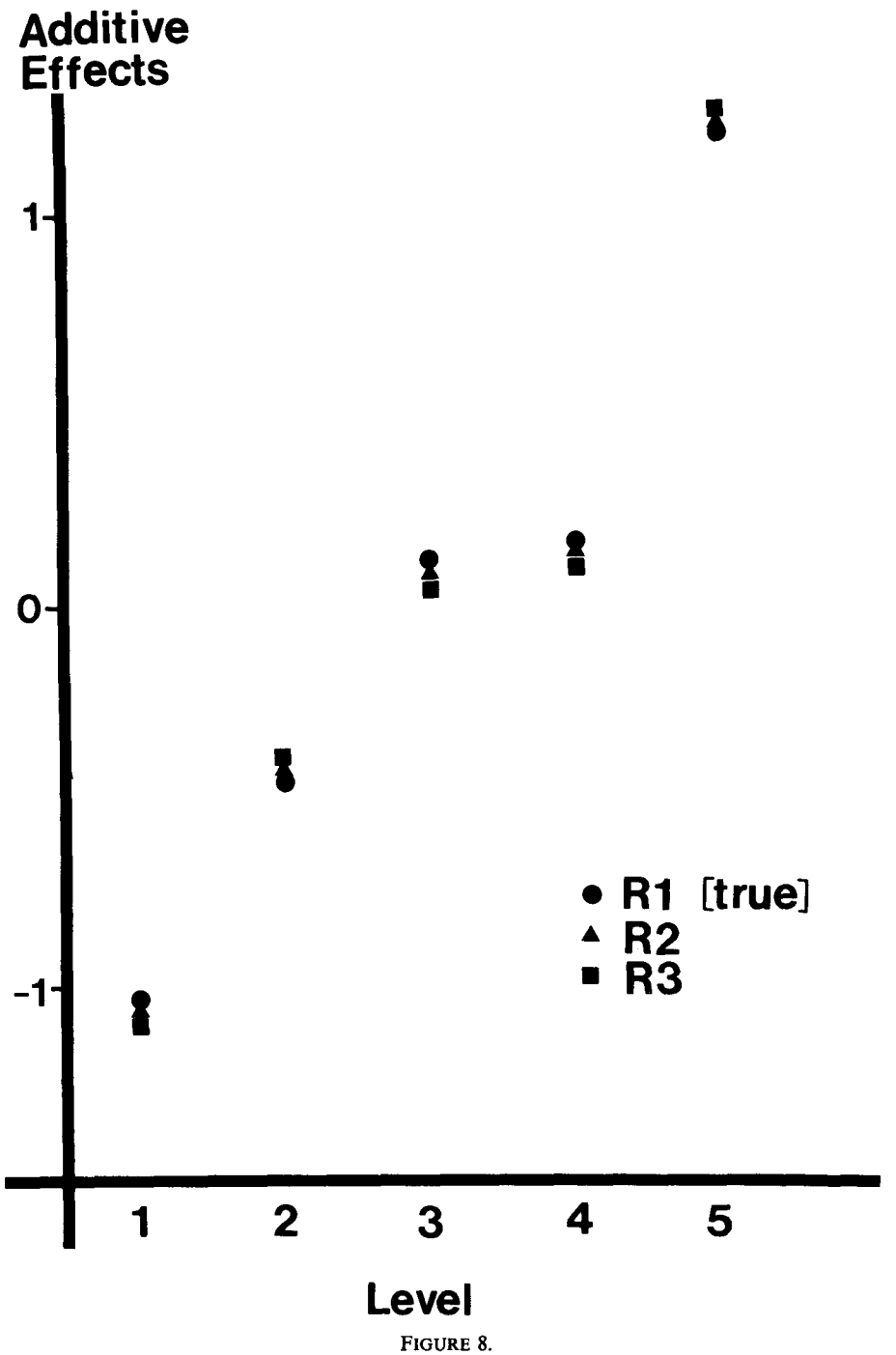

FIGURE 8.
Estimated additive effects of Factor B as a function of random error levels.

Model evaluations. The goodness of fit of the overall WADDALS analysis was .190 in the stress value (.204 for replication 1 and .173 for replication 2). Given that our results are stable we may compare this value with a variety of criteria to evaluate the goodness of fit of the model. One obvious criterion is that the observed data are completely random. We are typically interested in comparing the observed stress value with the expected stress

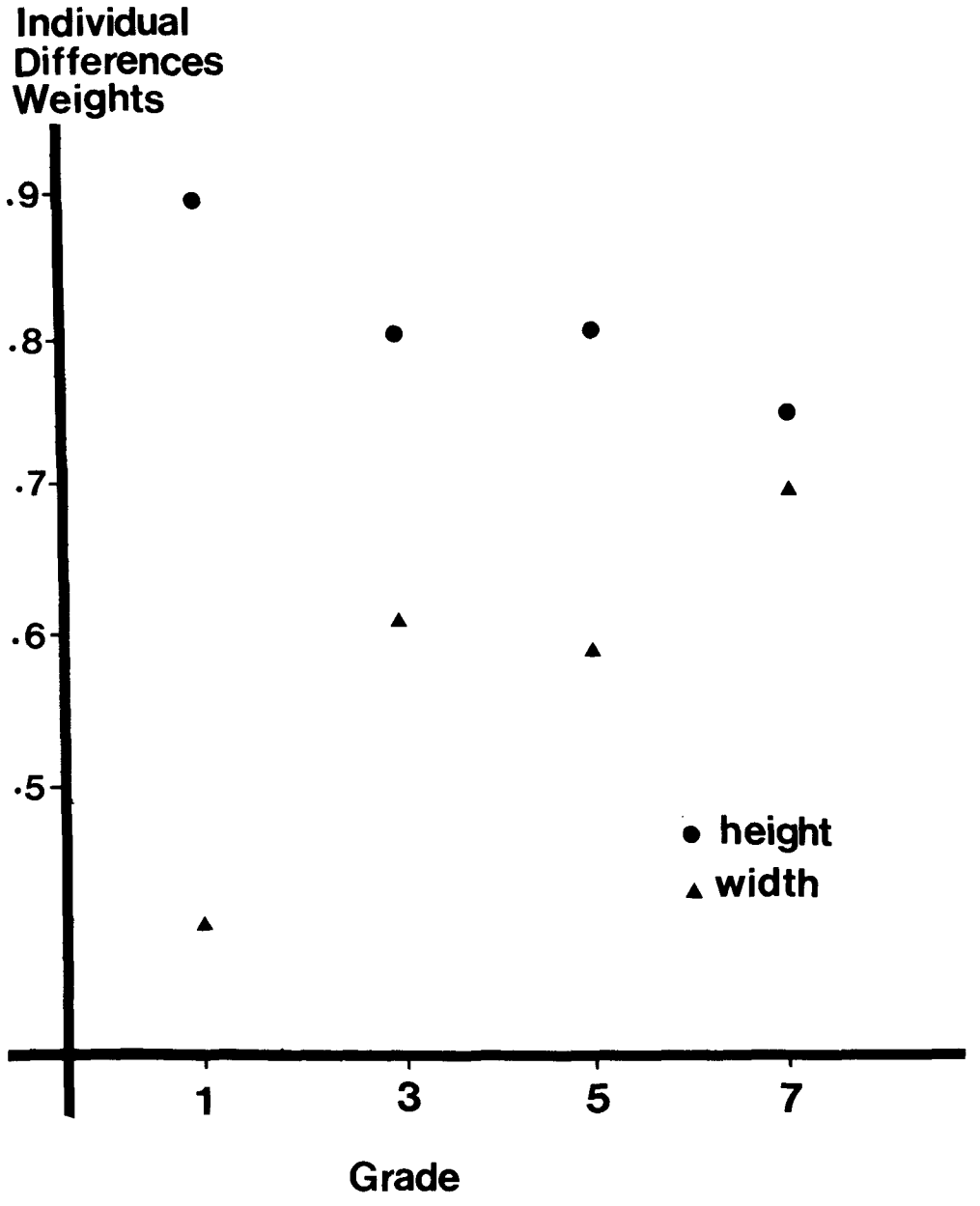

FIGURE 9.
Kempler data: Changes in weights attached to height and width of rectangles in area judgments as a function of age groups.

value from completely random data to make sure that we are not fitting the model to random data. An extensive random ranking study has been conducted to tabulate the expected stress values under various conditions. We have so far examined the two-factor case in which the numbers of levels in the two factors and the number of individuals were systematically varied (n_A, n_B = 4, 5, 6, 8 and 10, and N = 1, 2, 4, 6 and 10). The $n_A \times n_B \times N$ uniform random numbers were generated for each data set. Ten independent data sets were sampled in each condition.

The following formula has been derived from the empirical stress values obtained from the WADDALS analysis of these data:

$$s = .900 + .851 \times \{(\ln n_A)^{\frac{1}{4}} + (\ln n_B)^{\frac{1}{4}}\} + .895 \times 10^{-1} \times (\ln N)^{\frac{1}{3}} - .387 \times (\ln n_A \times \ln n_B)^{\frac{1}{4}}, \tag{21}$$

where s denotes a predicted stress value from random data as a function of n_A, n_B and N. The standard error of this regression was .045. According to this formula we expect to observe $s = .904$ under the condition equivalent to Kempler's data (i.e., when $n_A = n_B = 10$ and $N = 4$) if they were in fact completely random. (The observed value of expected s was .909 with the standard deviation of .006.) The stress value of .190 is significantly smaller than this value.

The next interesting criterion would be the assumption of individual-wise additivity. That is, each individual's data set is assumed to have a perfect simple additive representation, though the total data set may not have a weighted additive representation. This is clearly not the case with Kempler's data. (The stress values were substantially larger than zero when the simple additive model was separately fit to individual data sets.) Nonetheless, it is interesting to know what stress value is likely to come up when the perfect additivity assumption holds for each individual data set. We have not investigated this problem systematically, but have only obtained the mean stress value under the equivalent condition to Kempler's data. Twenty uniform random numbers were generated for $\mathbf{a}_{Ak}$ and $\mathbf{a}_{Bk}$ from which data were calculated according to the simple additive model for each individual. The mean stress value (obtained from ten replications) was .668 with the standard deviation of .034. Although this stress value is much smaller than that from completely random data, it is still substantially larger than the stress value found in Kempler's data, which supposedly contain a sizable amount of random errors (in addition to deviations from the weighted additive model).

Another interesting criterion is that the effect of age groups is not multiplicative, but rather additive. In order to test this hypothesis Kempler's data were analyzed by the three-factor (height, width and age-group) simple additive model. The obtained stress value was .269. By the procedure to be described below the standard error of the stress value around .190 from the weighted additive model is estimated to be about .025 under the equivalent condition. The difference of .079 is more than three times as large as this estimated standard error. Thus, it seems that the weighted additive model with the multiplicative age-group effect provides a better account of Kempler's data than the three-factor simple additive model.

On the basis of the same reasoning we may compare the joint stress value obtained from applying the simple additive model to individual data sets with that obtained from the weighted additive model. The difference of .018 (= .190 − .172) is well within the range of what can be expected by chance. Furthermore, the simple additive model separately fitted to individual data sets uses distinctly more parameters than the weighted additive model to describe the same set of data.

Our final concern relates to whether the weighted additive model is the best model conceivable for Kempler's data. Isn't there any interaction between height and width? How should we specifically interpret the stress value of .190? In an attempt to attack these problems we first tried to assess the magnitude of error variance relative to the magnitude of systematic variation (which is fixed at unity). Ten sets of data were generated for each value of $\sigma_e^2 = .01, .025, .05, .1, .25, .5$ and 1.0 in the same way as in the first Monte Carlo study (see section of the "goodness of recovery of the original information") under conditions equivalent to Kempler's. The mean stress value and the standard deviation were cal-

culated for each σ_e^2. In order to develop a formula which relates a stress value to σ_e^2, $\ln \sigma_e^2$ is fitted to ln s, the log stress value. The following regression equation was obtained:

$$\ln s = -.389 + .498 \ln \sigma_e^2 . \tag{22}$$

The standard error in this regression problem is about .009 which compares favorably with the standard error of .007 for the mean stress values. From this equation it is found that the stress value of .190 corresponds to $\sigma_e^2 = .078$. That is, systematic variation (the joint effect of height and width on the judgment of rectangle area) is about 14 times as large as σ_e^2. However, what proportion of σ_e^2 is due to the interaction is yet to be investigated.

For this purpose the data from two replicated experiments were separately analyzed by WADDALS. Following the functional measurement methodology [Anderson, 1977] the optimally scaled data obtained from WADDALS analyses were reanalyzed by ANOVA, separately for each age group. For these analyses two replicated experiments were treated as if they were independent samples. In all four age groups the interaction mean squares were 1.7 to 2.8 times as large as the lumped error mean squares. Taken nominally they indicate significant interaction effects.

The above result should be taken with some caution, however. (This, however, should not be taken as caution against the functional measurement methodology in general.) First, observations from the two replicated experiments were assumed to be independent, which was obviously not strictly true. It is likely that the error variance is drastically underestimated because of the correlated observations. Second, the optimally scaled data obtained from the same WADDALS analysis are also correlated, and possibly the degree of correlation is not homogeneous. The F-tests associated with the repeated-measures ANOVA tend to be too liberal in such situations. Third, we failed to observe any characteristic patterns of interaction effects both within and over different age groups. Thus, the best we could conclude at this point seems that the interaction effect is at most marginal, and its nature is yet to be investigated.

Discussion. The importance of a formula like (22) from a methodological point of view should not be overlooked. It tells us the variance component of the error term relative to the systematic variation, which in turn can be used in (19) and (20) to predict the goodness of recovery of the original information. Although (22) is restricted to the case in which we have $n_A = n_B = 10$ and $N = 4$, we should be able to derive similar equations under different conditions, and ultimately to combine them into a single equation. In general the variance component of the error term has some methodological advantage over stress, which is a function of many variables (such as the number of factors, the number of levels in each factor, and the number of individuals) which are irrelevant for the purpose of model evaluation. The error variance, on the other hand, can be determined, using the above techniques, independently of such irrelevant factors. We contend that, unless we are comparing the stress values obtained under equivalent conditions, the stress information should always be converted into the variance component term.

The above result obtained from Kempler's data (that the developmental change in children's perceptual structures of rectangle areas can be described reasonably well by the weighted additive model) apparently contradicts two recent findings on the same topic. Anderson and Cuneo [1978] and Wilkening [1979] argue that combination rules governing children's and adults' perceptual structures are entirely different; whereas adults judge the area of rectangles by the height × width rule, younger children typically judge the area by the height + width rule. Although both of these rules are subsumed under the additive conjoint measurement, they are not amenable to the weighted additive model when combined.

There are several conceivable reasons for this apparent discrepancy. First, the log transformation is so nearly linear within the range of height and width levels (10 inches to 14.5 inches) employed in Kempler's study that clear distinctions between the additive and multiplicative rules do not emerge. Estimated additive effects were in all cases roughly linear with respect to the physical length in Kempler's data. On the other hand our result might be due to some peculiarity of the judgment employed in Kempler's study. The simple two-category judgments have some advantage over rating judgments to ensure that the measuring device is used in exactly the same way across different age groups. Nonetheless the nature of the task, or more precisely the psychological processes mediating two-category judgments might be quite distinct from those involved in rating judgments. Still, a puzzling thing is why we rather consistently find larger weights attached to height (e.g., for Stimuli 1 and 2 with $\text{height}_1 = \text{width}_2 > \text{width}_1 = \text{height}_2$, we find $\text{area}_1 > \text{area}_2$ in Kempler's data), whereas the other investigators find approximately equal weights (e.g., for the same two stimuli given above, they find $\text{area}_1 = \text{area}_2$). Again this might be due to the difference in the judgmental tasks, although nothing definite can be said until Kempler's result is replicated using independent samples.

Conclusion

We have examined various aspects of the performance of the WADDALS procedure (the weighted additive model and the associated parameter estimation procedure based on the alternating least squares principle) through the analyses of empirical data. We conclude that it provides a useful approach to certain types of additivity problems in psychology. Note, however, that for model evaluation we still largely have to rely on Monte Carlo techniques (as in all other nonmetric procedures within Kruskal's transformational approach [Kruskal, 1964]), which sometimes makes rigorous hypothesis testing rather involved, even after discounting the fact that the WADDALS program has been written in such a way that the kinds of Monte Carlo experiments reported in this paper can be readily conducted by specifying just a few relevant parameters. One promising way to get out of this difficulty is to apply the maximum likelihood method for nonmetric multidimensional scaling suggested by Takane [1978, Note 2] to the present case, although currently this procedure imposes rather stringent restrictions on the type of data to be analyzed.

REFERENCE NOTES

1. de Leeuw, J. *Normalized cone regression approach to alternating least squares method.* Unpublished manuscript. The University of Leiden, 1977.
2. Takane, Y. *Statistical procedures for nonmetric multidimensional scaling.* Unpublished doctoral dissertation. The University of North Carolina, 1977.

REFERENCES

Anderson, N. H. Failure of additivity in bisection of length. *Perception & Psychophysics*, 1977, *22*, 213–222.

Anderson, H. H. & Cuneo, D. O. The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General*, 1978, *107*, 335–378.

Carroll, J. D. & Chang, J. J. Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 1970, *35*, 283–319.

Graybill, F. A. *Introduction to matrices with applications in statistics.* New York: Wadsworth, 1969.

Kempler, B. Stimulus correlates of area judgments: A psychophysical developmental study. *Developmental Psychology*, 1971, *4*, 158–163.

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

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964, *29*, 115–129.

Kruskal, J. B. Analysis of factorial experiments by estimating monotone transformations of data. *Journal of the Royal Statistical Society, Series B,* 1965, *27,* 251–265.

Lawson, C. L., & Hanson, R. J. *Solving least squares problems.* Englewood Cliffs, N.J.: Prentice Hall, 1974.

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

Liebert, R. N., Poulos, R. W. & Strauss, G. *Developmental psychology.* Englewood Cliffs, N.J.: Prentice Hall, 1974.

Nishisato, S. *Analysis of categorical data.* Toronto: University of Toronto Press, 1979 (in press).

Rao, C. R. *Linear statistical inference and its applications.* New York: Wiley, 1973 (1st edition: 1965).

Roskam, E. E. *Metric analysis of ordinal data in psychology.* Voorschoten, Netherlands, 1968.

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

Spence, I. A direct approximation for random rankings stress values. *Multivariate Behavioral Research,* 1979 (in press).

Takane, Y. 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,* 1978, *20,* 7–17 and 105–114.

Takane, Y., Young, F. W. & de Leeuw, J. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika,* 1977, *42,* 7–67.

Wilkening, F. Combining of stimulus dimensions in children's and adults' judgments of area: An information integration analysis. *Developmental Psychology,* 1979, *15,* 25–33.

Young, F. W., de Leeuw, J. & Takane, Y. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika,* 1976, *41,* 505–529.

Young, F. W., de Leeuw, J. & Takane, Y. *Quantifying qualitative data.* In H. Feger (Ed.). *Similarity and choice.* New York: Academic Press, 1979.

Zangwill, W. I. *Nonlinear programming: A unified approach.* Englewood Cliffs, N.J.: Prentice Hall, 1969.

Manuscript received 12/13/77
First revision received 6/11/79
Final version received 11/28/79