T_1 SCALING

A Mathematical Programming Approach
to Thurstonian Scaling

James M. Lattin

T_1 SCALING

A Mathematical Programming Approach
to Thurstonian Scaling

James M. Lattin

WP 1183-81   January 1981

ABSTRACT

We present a new technique, called T_1 scaling, for
determining scale estimates from paired comparisons data. We
present the new method in conjunction with a sensitivity
diagnostic that ascertains the extent to which intransitive
elements in the data influence the scale estimates from the
Thurstonian judgment scaling model. The T_1 scale estimates,
based upon the minimization of absolute deviations rather than
least squares, are relatively insensitive to the presence of
limited inconsistency. We apply the new solution technique,
shown to be a straightforward minimum cost network flow problem,
to several scaling problems in the literature. When no single
limited source of inconsistency is indicated, the scale estimates
thus obtained are consistent with the least squares estimates.
When isolated departures from the scaling model or possible data
errors are present, the T_1 procedure remains largely insensitive
to their presence, preserving the interval scale properties of
the estimates.

74006

Introduction

This paper presents a new solution technique, called T_1 scaling, for determining scale estimates from paired comparisons data. The development of the T_1 procedure was motivated by a concern for the substantial influence of intransitive elements in the data on the final solution produced by the least squares approach of the Thurstonian judgment scaling model (Thurstone, 1927a). The new technique utilizes a discrete L_1 linear approximation based upon the minimization of absolute deviations (see Barrodale, 1968, and Barrodale and Roberts, 1973), where the special structure of the scaling problem allows it to be solved quite efficiently as a minimum cost network flow problem, using standard techniques presented in such sources as Bradley, Hax, and Magnanti (1977) or Shapiro (1979). The scale estimates thus provided are in some sense more "robust" than in the traditional approach in that they discount the influence of limited inconsistency in the data.

The balance of the paper follows in several sections. The first section reviews the least squares solution technique for the Thurstonian judgment scaling model. The second section reviews Mosteller's goodness-of-fit measure for the least squares estimates (Mosteller, 1951), and demonstrates how seriously this fit deteriorates in the presence of limited inconsistency. In order to do this, we develop a sensitivity diagnostic to assess the relative influence of each pair of items on the determination

2

of the least squares estimates. The third section presents the
robust T_1 scaling technique and shows that obtaining these scale
estimates is equivalent to solving a minimum cost network flow
problem. Finally, we apply the T_1 approach to several problems
from the literature, and compare the results to the least squares
scale estimates.

I. The Law of Comparative Judgment and the Thurstonian Judgment
Scaling Model

    In the typical judgment scaling problem, we are presented
with k different objects, each exhibiting some degree of a
certain common characteristic. If this characteristic, such as
"height", "weight", or "age", is a directly measureable quality
of singular dimensionality, then we can order these k objects by
placing them along a continuum at the measured value of their
common characteristic. The positions of these objects, or scale
values, have the properties of the measurement scale. For
example, objects ordered on the basis of height or weight have
scale values with ratio properties, while objects ordered on the
basis of heat in degrees Farenheit have interval properties.

        When the objects share a common characteristic that is
not directly measurable, such as "beauty" or "softness", the
ordering of the objects must depend upon some subjective estimate
of the common characteristic exhibited by each object. In order
to facilitate the process of ordering the objects along a
continuum without an apparent scale, the method of paired
comparisons is used to exact a set of relative judgments from an
observer. Thus, for any given pair of objects the observer is
required only to judge which of the two exceeds the other with
respect to the underlying characteristic. This set of pairwise
judgments is used to determine scale values with interval
properties.

        The law of comparative judgment established the

theoretical foundations for Thurstone's judgment scaling model. Each object, when presented to the observer, acts as a stimulus which excites a certain discriminal process within the observer. Due to changing conditions in the experimental situation or fluctuations within the observer, the same stimulus might trigger a slightly different process, such that the position of the stimulus on the specific psychological continuum is not always the same. For example, an observer's subjective estimate of the "beauty" of an object might be different when presented with the object a second time, on account of the observer's mood, the time of day, or the temperature of his surroundings.

In the Thurstonian model, the distribution of these subjective estimates along the continuum is postulated to be normal. The standard deviation of this distribution is called the discriminal dispersion of the stimulus, and the mean is taken to be the true scale value. The distributions of two stimuli, i and j, might thus be represented as in Figure I.1. The scale values are s_i and s_j, and the discriminal dispersions are the standard deviations, σ_i and σ_j. The discriminal processed within the observer are random variables denoted d_i and d_j.

It is now possible to talk about a discriminal difference, $(d_i - d_j)$, for any pair of stimuli i and j. If i and j are presented to an observer a large number of times, the discriminal differences will also form a normal distribution, with standard deviation $\sigma_d = (\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j)^{1/2}$, where r_{ij} is the correlation between the discriminal processes associated with i and j.
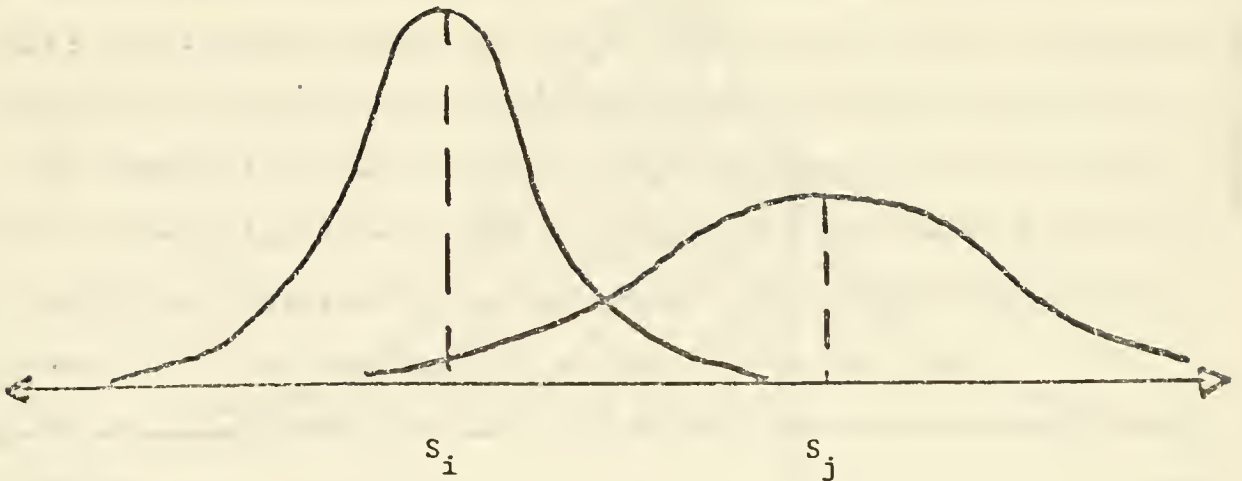
5

FIGURE I.1

The discriminal distributions for stimuli i and j,
centered about the true scale values s_i and s_j.

The observer, of course, is unable to assign a value to

the position of the stimulus along the appropriate psychological

continuum, but when presented with two stimuli, he is able to

judge which of the two is greater. In some cases, because the

distributions overlap, the observer may judge stimulus i to be

greater than j even though s_j is actually greater than s_i. Over

a large number of comparisons, it is possible to determine the

approximate proportions of times stimulus j is judged greater

than i. These proportions are then used to determine the

relative positions of s_i and s_j on the continuum measuring their

common quality.

Figure I.2 shows the distribution of the discriminal

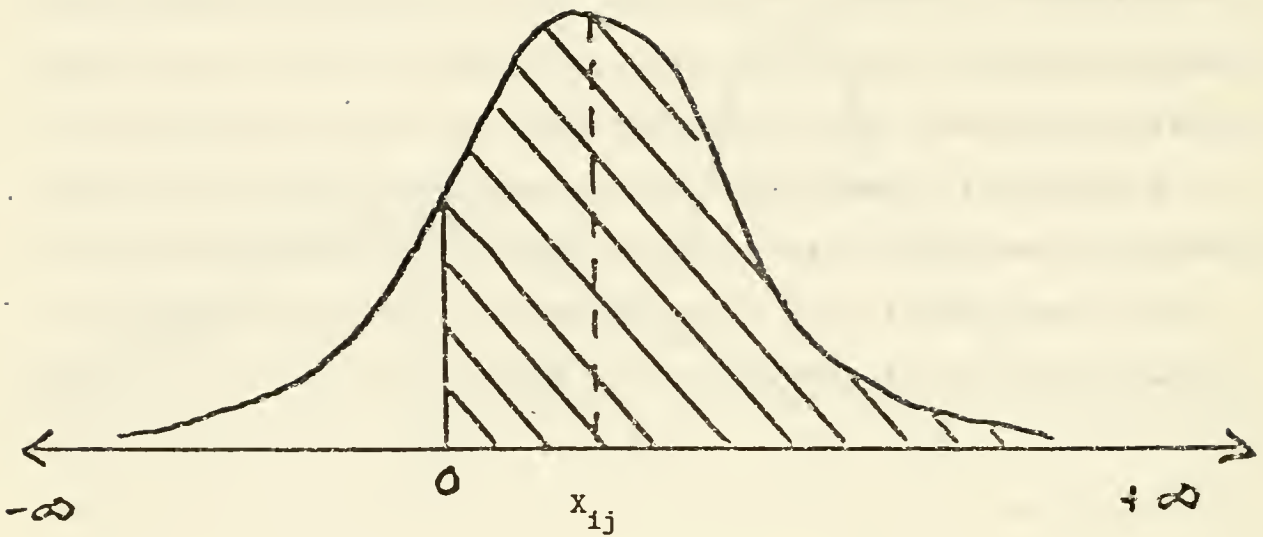difference between i and j, where the shaded portion of the curve

FIGURE I.2

The discriminal difference between stimuli i and j.

indicates the proportion of times stimulus j appears greater than stimulus i to the observer. The value x_{ij} is the difference between s_i and s_j measured in σ_d units. Hence, $s_j - s_i = x_{ij}\sigma_d$, or in its final form:

$$(1) \qquad s_j - s_i = x_{ij} (\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j)^{1/2}$$

In this form, without limiting assumptions, the law of comparative judgment is not solvable, as there are many more unknowns than equations. Thurstone presented five cases of the law, introducing certain assumptions into the model to make it tractable. His case V is the most restrictive, assuming constant standard deviations for all of the discriminal dispersions, and

7

no correlation between any of the discriminal processes (implying zero covariance between stimuli). Mosteller (1951) later showed that the assumption of equal correlations between processes leads to a formulation equivalent to Thurstone's case V. In either case, the unit of measurement for the psychological scale may be determined arbitrarily; hence, the constant modifying the x_{ij} term in (1) may be taken to be unity, leaving

(2) $\qquad s_j - s_i = x_{ij}$

The law of comparative judgment, limited by assumptions of equal dispersions and equal covariances, is most frequently estimated using paired comparisons data. In this procedure, we present each pair of stimuli to the observer a large number of times, as described above. If we wish to examine the collective discriminal process of an entire population, we present each stimulus pair i,j to each individual in the population only once. Thurstone, in his case II of the law of comparative judgment, showed that the same formulation holds true for either approach, under certain assumptions of homogeneity.

The observed proportion of times stimulus j exceeds i, p'_{ij}, forms the matrix P'. Matrix P' has the property that symmetric cells must sum to one; hence, $p'_{ij} + p'_{ji} = 1$. Matrix P' determines matrix X', where each element x'_{ij} is the unit normal deviate corresponding to the observed proportion p'_{ij}. If the range of stimuli along the psychological continuum is large relative to the discriminal dispersion, there may in fact be

8

cases where one stimulus is never judged greater than another. If stimulus i is judged less than j every time the pair is presented, the value p'_{ij} will equal one and the value x'_{ij} will approach infinity. This problem of an incomplete X' matrix is usually solved by establishing upper and lower bounds on p'_{ij} of .01 and .99, thus insuring stability of the resulting scale estimates.

From (2) above, we see that the difference between the estimates of any two scale values, s'_i and s'_j, gives us x''_{ij}, an estimate of the observed value x'_{ij}, as shown:

(3) $$s'_j - s'_i \equiv x''_{ij}$$

With errorless data, we can choose scale estimates s'_i and s'_j so that the estimates x''_{ij} will equal the observed x'_{ij}. Typically, differences between the observed proportions and the true values lead to a difference between x''_{ij} and x'_{ij}, no matter how we choose s'_i and s'_j. Thurstone chose his scale estimates to minimize Q, the sum of the squared deviations between x''_{ij} and x'_{ij}:

(4) $$Q = \underset{ij}{\Sigma\Sigma} \; (x'_{ij} - x''_{ij})^2$$

Substituting (3) into the equation above:

(5) $$Q = \sum_{ij} (x'_{ij} - s'_j + s'_i)^2$$

Equation (5) is equivalent to minimizing either row sums or column sums, so Thurstone limited his analysis to the columns of X'.

Differentiating Q with respect to s'_j gives:

(6) $$\frac{dQ}{ds'_j} = -2 \sum_i (x'_{ij} - s'_j + s'_i)$$

Setting the partial derivative to zero and solving :

(7) $$s'_j = \frac{1}{k} \sum_i x'_{ij} + \frac{1}{k} \sum_i s'_i$$

where k is the number of objects in the scaling problem. The rightmost term in (7) is simply the mean of the estimated scale values. Because the origin of the psychological continuum is arbitrary, we can take it to be the mean of the s'_i, giving:

(8) $$s'_j = \frac{1}{k} \sum_i x'_{ij}$$

Thus, the least squares estimates of the true scale values are the column means of the matrix X'. Torgerson (1958) presents a more detailed discussion of this derivation.

II. The Effects of Inconsistency in the Observed Data on the

Determination of the Thurstonian Scale Estimates

The method of paired comparisons does not always produce a
set of data appropriate for use in a scaling model such as
Thurstone's. If there is perfect agreement among the judges on
the ordering of the k objects being compared, then it is not
possible to determine scale estimates with interval
characteristics. Another problem occurs when an observer or some
observers are particulary bad judges, or are poorly motivated to
take the care required to produce consistent comparisons. A
third problem occurs if the experimenter asks too much of his
observers; the objects may be so close together with respect to
their common quality that distinguishing them becomes almost a
guessing game. Finally, it is possible that the quality common
to the objects under examination is not representable as a linear
variate. When any one or several of these difficulties is
present, the reported preferences may contain intransitivities
called circular triads, where object i is judged greater than
object j, j is judged greater than k, yet k is ultimately judged
greater than i. Such an ordering is impossible to represent on a
single dimensional scale, and thus interferes with the process of
determining scale estimates.

Kendall and Babington Smith (1940) observed that
"[Thurstone's] method is appropriate where one is entitled to
assume _a priori_ or by reason of precautions taken in the

11

selection of material that a linear variable is involved and that there exist perceptible differences between the items presented for comparison."(p. 342) They proposed a coefficient of consistence, ζ , where

$$\zeta = \begin{cases} 1 - 24d/(k^3 - k) & \text{k odd} \\ 1 - 24d/(k^3 - 4k) & \text{k even,} \end{cases}$$

where k is the number of objects and where d is the number of circular triads reported. The coefficient equals one when the comparisons data contain no inconsistencies and equals zero when the maximum number of circular triads is present. Thus, a value of ζ near zero indicates potentially troublesome departures from the scaling model.

For paired comparisons with fewer than eight objects, Kendall and Smith also calculated the probabilities that a number of circular triads d or greater would occur under a completely random ranking scheme. If a single observer reports a number of circular triads d that is likely to have come from a process of unsystematic (random) judgment, his ability to discriminate between objects should be questioned; if a number of observers do the same, then a problem may lie in the difficulty of the task or in the dimensionality of the quality under judgment.

Even when paired comparisons data are free of complete intransitivity, there is usually some form of inconsistency present. In Figure II.1 below, three stimuli a, b, and c are shown equally spaced along the appropriate psychological

continuum. By assumption, their discriminal dispersions are all equal (Thurstone's case V), and because the origin for the scale is arbitrary it has been placed at the middle scale value. If we presented an observer with stimulus pair a,b and stimulus pair b,c a total of n times each, it is unlikely (due to statistical fluctuation) that the observer would report a>b exactly the same number of times he reported b>c. Even so, while an observer may judge a>b and b>c approximately the same number of times each, he might judge a>c only a slightly higher number of times, not necessarily consistent with placing c twice as far from a as from b.
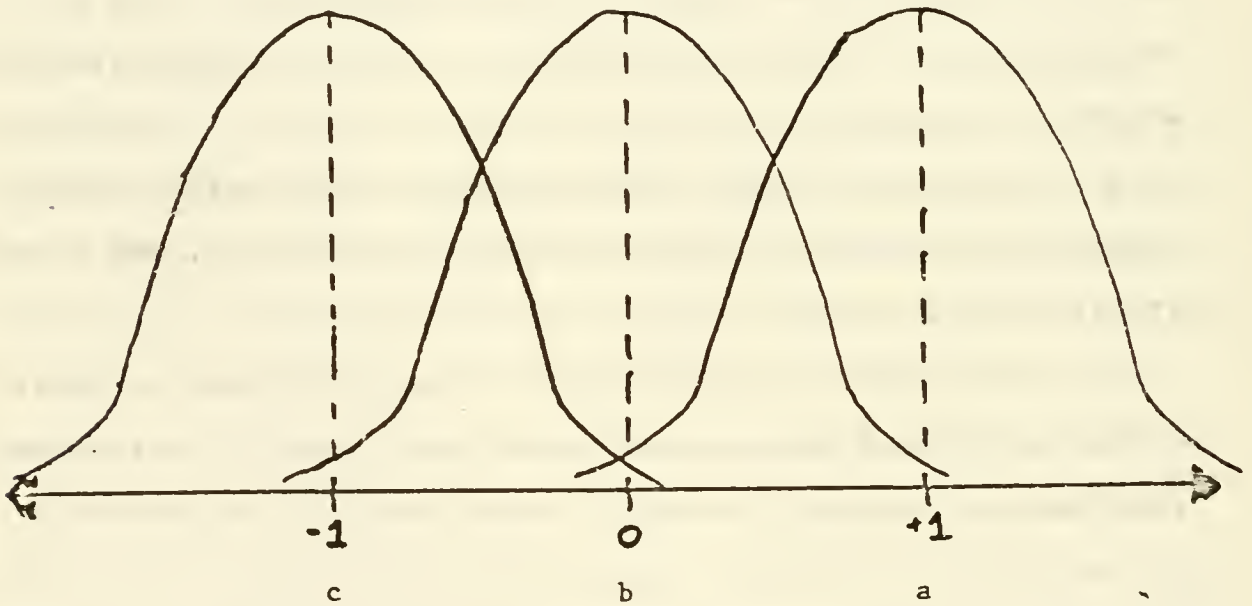

FIGURE II.1

True underlying model for the three stimulus example demonstrating that inconsistency need not take the form of intransitivity.

13

With Thurstonian scale estimates, stimulus c is positioned on the scale somewhat closer to b than suggested by the proportion of times b>c, yet not so close to a as suggested by the proportion of times a>c. This "compromise" leads to discrepancies between the observed values x'_{ij} and the values x''_{ij} derived from the scale estimates, leaving a question as to the fit of the final result.

With fallible data, it is helpful to have a measure of the goodness-of-fit of the least squares estimates. Mosteller (1951) presented a chi-square significance test for the fit between the observed proportions p'_{ij} and the fitted proportions p''_{ij}. These fitted proportions are derived from the scale estimates; they represent the proportion of the time stimulus i would be judged greater than stimulus j if the true scale values were actually s'_i and s'_j. We can use the unit normal table to find the proportion p''_{ij} corresponding to each x''_{ij}, and form the matrix of fitted proportions P".

Mosteller suggested the arcsin transformation developed by R. A. Fisher to establish a chi-square testing criterion. Given proportions p'_{ij} and p''_{ij} from a binomial sample of size n,

$$\theta' = \arcsin \sqrt{p'} \quad \text{and} \quad \theta'' = \arcsin \sqrt{p''}$$

are distributed with variance

$$\sigma^2 = \frac{821}{n}$$

14

when $\theta'ij$ and $\theta''ij$ are expressed in degrees. Thus, Mosteller suggests the following test of the goodness-of-fit of the estimates:

$$\chi^2 = \sum_{i>j} \frac{(\theta''_{ij} - \theta'_{ij})^2}{821/n}$$

where n is the total number of times each stimulus pair is presented. The test covers the elements in the lower triangular matrix; thus, for a scaling problem involving k stimuli, the distribution is $\sim \chi^2 ((k-1)(k-2)/2)$.

It is possible to assess the nature of the effect of inconsistency on the fit of the scale estimates by constructing a situation in which a single circular triad is exhibited in otherwise errorless comparisons data. Figure II.2 shows the placement of four stimuli, a, b, d, and e, along the psychological continuum. The differences between these actual scale values are shown in the matrix X' in Table II.1. The fifth stimulus, c, is represented at two positions on the scale. With respect to all stimuli but a, c is positioned at -.10 on the scale (the true value, c_{bde}). With respect to a, however, c is positioned at +.15 on the continuum (c_a). The result is a single inaccurate observation for stimulus pair a,c, forming the single circular triad (a>b, b>c, c>a).

If the observed proportion of times that c was judged greater than a were overlooked, perhaps discounted as a transcription error, the remaining data in X' would be consistent
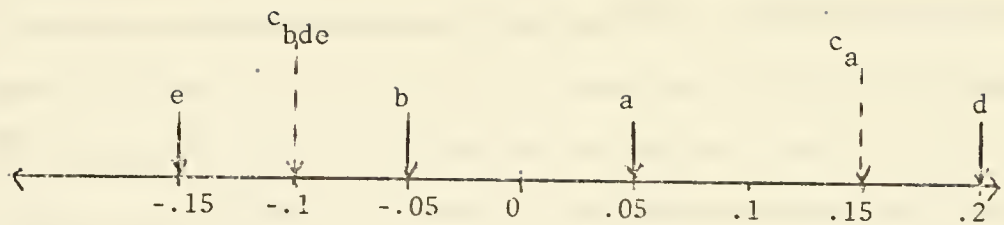
15

FIGURE II.2

Contrived five-stimulus example, where inaccuracy is introduced
into the observation between stimuli a and c.

	a	b	c	d	e
a	–	-.1	.1	.15	-.20
b	.1	–	-.05	.25	-.1
c	-.1	.15	–	.35	-.05
d	-.15	-.25	-.35	–	-.35
e	.20	.1	.05	.35	–
$\underline{s}' =$	.01	-.04	-.04	.21	-.14

TABLE II.1

Matrix X' for the contrived five-
stimulus example of Figure II.2.

with the determination of scale estimates leading to a perfect fit between P" and P'. Including the inconsistency produces the distorted scale estimates shown in Figure II.3. Only the scale estimates for stimuli a and c are different from the true underlying values; stimuli b, d, and have the same relative positions as in the actual configuration. Because X' is a skew symmetric matrix ($x'_{ac} = -x'_{ca}$) and the only columns affected by the presence of intransitivity are those for a and c; the estimated scale values differ from the actual values by the same amount in opposite direcitons. As shown in Figure II.3, the scale estimate for stimulus c is .05 units greater than its actual value; for a, it is .05 units less.
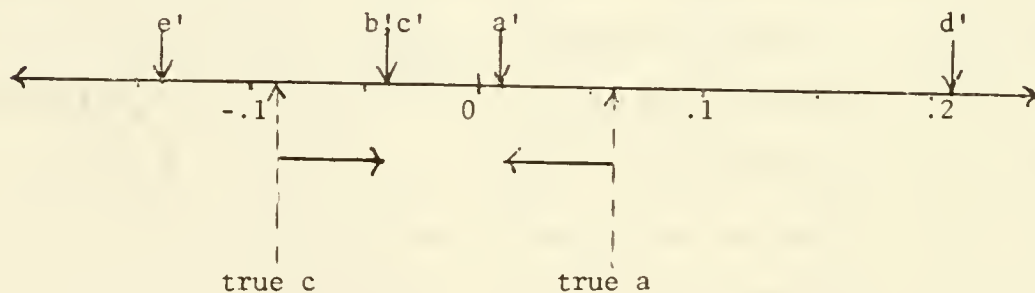

FIGURE II.3

Thurstonian scale estimates for the five-stimulus example compared to the true values for stimuli a and c.

The distortion introduced into the scale due to the inaccurate comparison of a and c degrades the fit of the least squares estimates noticeably. In this example, where a single

circular triad is introduced into otherwise errorless data, the
fitted proportions differ from the observed in seven out of 10
cases, as shown in Table II.2. Because the least squares
procedure operates to minimize the sum of squared deviations, a
solution resulting in several small discrepancies is preferred to
a solution with a single large one. Thus, the least squares
procedure distorts the interval properties of several scale
estimates in order to compensate for a single potentially
problematic observation.

	a	b	c	d	e
a	–				
	–				
b	.54	–			
	.516	–			
c	.46	.54	–		
	.532	.516	–		
d	.44	.401	.363	--	
	.417	.401	.386	–	
e	.56	.52	.48	.618	–
	.536	.52	.504	.618	–

$$\left(\begin{array}{c} P'_{ij} \\ P''_{ij} \end{array} \right)$$

TABLE II.2

Comparison of observed and fitted proportions for the five-stimulus
example, showing discrepancey in seven out of 10 cells in the
lower diagonal matrix.

This small ad hoc analysis of limited intransitivity

motivates the design of a more general procedure. It would be

advantageous to have a diagnostic to determine the influence of

any one observation on the overall fit of the model. By

replacing the value in each cell of the lower diagonal of the

matrix X' by a value determined from the other relative

comparisons, and then assessing the fit for these modified

values, it is possible to determine the improvement in fit

associated with the "discounting" of one observation. If this

improvement is substantial, it indicates that the interval

properties of the initial scale may have been degraded by

inconsistency. If the inconsistency can be traced to data

transcription error, or to some other uncontrolled influence

operating on a limited portion of the data, we might want to turn

to a more robust scaling procedure where outlying observations

have less influence and the discrepancy in fit is limited to as

few stimulus pairs as possible.

Consider the contrived five-stimulus example shown in

Table II.1. In this case, we introduced an intransitivity by

perturbing the observed value x'_{ac}. If we could somehow discount

this observation, so that scale estimates s'_a and s'_c depended

only on the relative comparison with stimuli b, d, and e, the

resulting value would reflect the proportion of times stimulus a

was reported greater than c, with c positioned accurately. We

can thus use the concept of adjusting a stimulus pair to design a

diagnostic technique for determining the sensitivity of paired

comparisons data to inconsistency. This notion of sensitivity is

19

similar to the one introduced by Hoaglin and Welsch (1978), based
loosely on the influence of a single observation on the fit of
the entire model rather than simply its own fitted value.

If stimulus pair i,j is discounted, only one value in
each of column i and column j of matrix X' changes; therefore,
all other scale estimates s'_m, $m \neq i$ or j, remain unaltered. We
can use these k-2 unaffected scale estimates to adjust the values
s'_i and s'_j. The adjusted estimate $\hat{s}_i'$ will reflect the best
position for stimulus i relative to all other stimuli by j.
Similarly, $\hat{s}_i'$ will reflect the best position for stimulus j
without considering direct comparison to stimulus i.

The following sets of equations determine the adjusted
scale estimates $\hat{s}_i'$ and $\hat{s}_j'$:

$$
\begin{array}{ll}
\hat{s}_i' - s_1' = x_{1i}' & \hat{s}_j' - s_1' = x_{1j}' \\
\quad \vdots & \quad \vdots \\
\hat{s}_i' - s_{i-1}' = x_{i-1,i}' & \hat{s}_j' - s_{i-1}' = x_{i-1,j}' \\
\hat{s}_i' - s_{i+1}' = x_{i+1,i}' & \hat{s}_j' - s_{i+1}' = x_{i+1,j}' \\
\quad \vdots & \quad \vdots \\
\hat{s}_i' - s_{j-1}' = x_{j-1,i}' & \hat{s}_j' - s_{j-1}' = x_{j-1,j}' \\
\hat{s}_i' - s_{j+1}' = x_{j+1,i}' & \hat{s}_j' - s_{j+1}' = x_{j+1,j}' \\
\quad \vdots & \quad \vdots \\
\hat{s}_i' - s_k' = x_{k,i}' & \hat{s}_j' - s_k' = x_{k,j}'
\end{array}
\right\} \; \substack{k-2 \\ \text{equations}}
$$

With infallible data, such as that found in the five-stimulus example in Table II.1 above, all k-2 equations render exactly the same value for the adjusted scale estimate. However, since paired comparisons data rarely offer perfect observations, we again choose to use a least squares approach to obtain values for $\hat{s}_i'$ and $\hat{s}_j'$.

Our results above show that the mean of the k-2 equations gives the least squares solution:

$$\hat{s}_i' = \frac{1}{k-2} \sum_{\substack{m=1 \\ m \neq i,j}}^{k} (x_{m,i}' + s_m')$$

$$\hat{s}_j' = \frac{1}{k-2} \sum_{\substack{m=1 \\ m \neq i,j}}^{k} (x_{m,j}' + s_m')$$

Rearranging terms for $\hat{s}_i'$ gives:

$$\hat{s}_i' = \frac{1}{k-2} \sum_{\substack{m=1 \\ m \neq i,j}}^{k} x_{m,i}' + \frac{1}{k-2} \sum_{\substack{m=1 \\ m \neq i,j}}^{k} s_m'$$

Because the scale origin has been arbitrarily centered at the mean of the scale estimates, the equation above becomes:

$$\hat{s}'_i = \frac{1}{k-2} \sum_{\substack{m=1 \\ m \neq i,j}}^{k} x'_{m,i} + \frac{1}{k-2}(-s'_i - s'_j)$$

Similarly, because s'_i is equal to the mean of column i of X', the equation above becomes:

$$\hat{s}'_i = \frac{1}{k-2}(ks'_i - x'_{ji}) + \frac{1}{k-2}(-s'_i - s'_j)$$

With some rearrangement of terms, the adjusted scale estimates $\hat{s}'_i$ and $\hat{s}'_j$ can be written:

(9)
$$\hat{s}'_i = \frac{k-1}{k-2}s'_i - \frac{(x'_{ji} + s'_j)}{k-2}$$

(10)
$$\hat{s}'_j = \frac{k-1}{k-2}s'_j - \frac{(x'_{ij} + s'_i)}{k-2}$$

Examination of (9) and (10) reveals that $(\hat{s}'_i - s'_i) = -(\hat{s}'_j - s'_j)$. Thus, the adjusted estimates satisfy the symmetry exhibited in the contrived five-stimulus example, where the scale estimates for stimuli a and c moved the same distance in the opposite directions from their true scale value.

The adjusted scale estimates now uniquely determine new values for x''_{ij} and x''_{ij}. To assess the change in fit associated with adjusting the scale estimates for stimuli i and j, we form the adjusted matrices X" and P", denoted X"(i,j) and P"(i,j), and

22

use Mosteller's chi-square test with $((k-1)(k-2)/2 - 1)$ degrees of freedom.

We applied the diagnostic procedure to the contrived five-stimulus example from Table II.1. For each stimulus pair i,j, we adjusted the values of s'_i and s'_j and calculated the change in fit. The results showed little or no improvement in fit for all but the stimulus pair a,c. For that pair, the adjusted scale estimates for a and c equaled the true scale values for these stimuli, eliminating the source of intransitivity in the otherwise errorless model and indicating a complete improvement in fit.

In general, the diagnostic serves to identify sources of limited inconsistency or intransitivity in the paired comparisons data. If the data are widely inconsistent, then several of the scale estimates are liable to depart significantly from the true scale values. Using the diagnostic to adjust the scale estimates for a single stimulus pair might eliminate the inconsistency introduced by that particular observation, but the adjusted estimates would still reflect the inconsistencies that influenced the positioning of the other stimuli. Such widespread inconsistency, while not readily detectable by the diagnostic, usually shows up in a poor overall goodness-of-fit, indicating a departure from the assumptions made for the one dimensional scaling model.

Kendall and Babington Smith's coefficient of consistency is a valuable tool for identifying failure by a single observer to adequately discriminate between stimuli.

However, in Thurstone's case II of the law of comparative judgment, where the responses of several judges are used to determine the observed proportions p'_{ij}, all of the observers may give completely consistent responses, and yet the composite comparisons may contain inconsistency or even complete intransitivity, as shown in the example in Figure II.4. Our diagnostic functions as a computationally inexpensive indicator of limited inconsistency that is potentially damaging to the interval properties of the scale.

Judges 1,4	Judges 2,5	Judges 3,6	Composite	
A > B	C > A	B > C	A > B	(66%)
B > C	A > B	C > A	B > C	(66%)
A > C	C > B	B > A	C > A	(66%)

FIGURE II.4

An example demonstrating that the judgments of perfectly consistent observers may yield a perfectly intransitive composite ordering.

III. The T_1 Scaling Solution Procedure

Our concern for obtaining scale estimates that are
relatively insensitive to the presence of limited error or
intransitivity in the observed data motivates the development of
a more robust scaling procedure. The weakness with the
Thurstonian jugment scaling model in the presence of limited
inconsistency is that it is based on an L_2 linear approximation
of the underlying true values. In this least squares approach,
outlying observations tend to have an inordinate amount of
influence in the determination of the scale estimates. Because
least squares "prefers" a solution with several small
discrepancies to one with a single large error, the Thurstonian
procedure propagates limited inconsistency throughout the scale
estimates, degrading the interval properties of the entire scale.

Barrodale and Roberts (1973) suggest that when the data
contain inaccuracies or inconsistencies, an L_1 approximation,
minimizing the sum of the absolute deviations, is often superior
to the best L_2 approximation for estimating the true parameters
of the model. Thus, the L_1 approach to determining scale
estimates requires minimizing the quantity Q_i, where

(11) $$Q_1 = \sum_{i=1}^{k} \sum_{\substack{i=1 \\ j \neq i}}^{k} \left| x'_{ij} - (s'_j - s'_i) \right|$$

25

An example taken from simple regression, shown below in Figure III.1, illustrates their point. In the example, the true underlying values follow exactly a linear model. Only one of the seven observations differs from its true value, but that difference is quite substantial. The L_2 approximation operates to distribute this error across all seven points, and thus the seventh observation has the effect of tilting down the slope of the regression line to b' and raising the intercept to a'. The L_1 approximation is not so influenced by the seventh observation, and recovers the true model parameters, a and b.
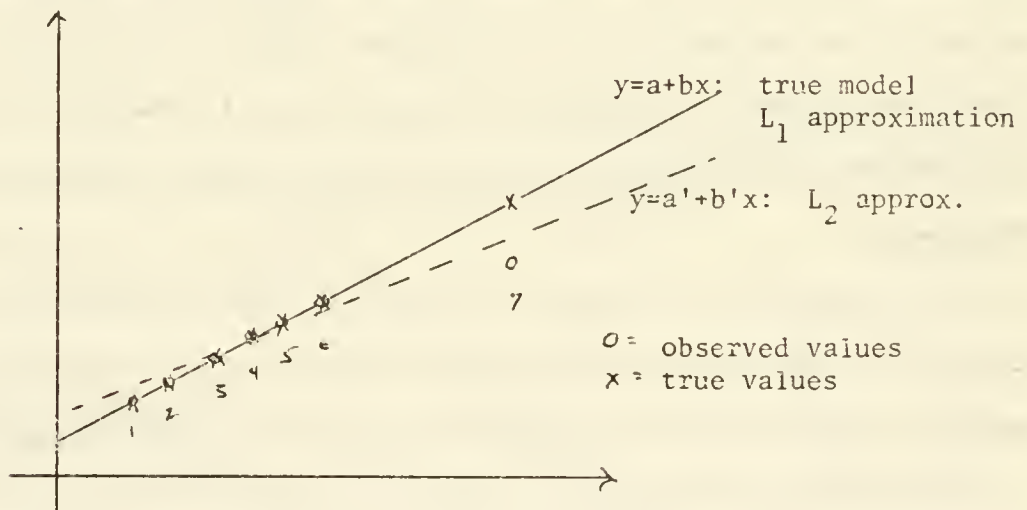
FIGURE III.1

Regression example demonstrating the relative insensitivity of the L_1 approximation to outlying observations.[1]

There are also weaknesses to the L_1 approach that should not be overlooked. The L_1 procedure occasionally performs very badly in the presence of an outlying observation, as shown in Figure III.2. In this example, the seventh observation deviates so substantially from its true value that a better fit is obtained by approximating the model using only the first and seventh points rather than using the first six. A second weakness of the L_1 technique with respect to the scaling problem at hand is that it has far greater computational requirements than simply using column means to estimate the scale values. Therefore, it remains for us to show that the L_1 approach applied to Thurstonian scaling, henceforth denoted T_1 scaling, can be solved in a manner that is computationally convenient, and that under reasonable assumptions regarding the behavior of the data the T_1 scale estimates are superior to those determined by least squares.
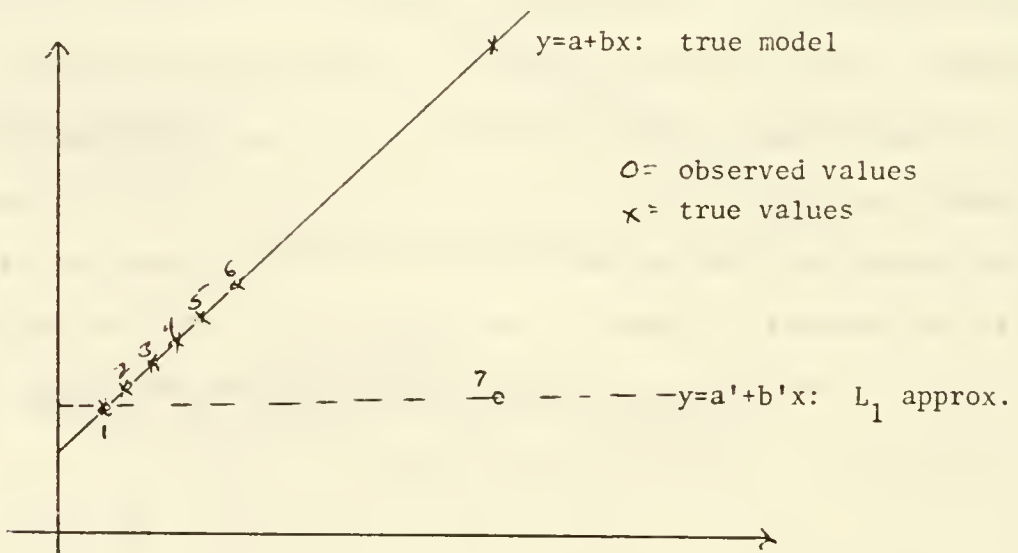
FIGURE III.2

Regression example demonstrating the weakness of L_1 approximation in the presence of a wildly inaccurate observation.

We now show that solving the T_1 scaling problem for k scale estimates is equivalent to solving a capacitated network flow problem for a complete network with k nodes (see Bradley, Hax, and Magnanti for a description of this type of problem). Because X' is skew symmetric, we may limit our analysis to the lower diagonal matrix, writing (11) as follows:

$$(12) \qquad Q_1 = \sum_{i>j} \left| x'_{ij} - (s'_j - s'_i) \right|$$

Using standard techniques, the minimization problem described above may be written as a linear program:

$$\text{Minimize} \quad \sum_{m=1}^{k(k-1)/2} (u_m + v_m)$$

(13) $\qquad$ subject to $\quad s'_j - s'_i + u_m - v_m = x'_{ij} \quad$ for all (i,j) st $i>j$

$$m = 1, 2, \ldots, k(k-1)/2$$

$$u_m, v_m \geq 0, \; s'_j \text{ unconstrained}$$

Thus, for a scaling problem with k stimuli, there are $k(k-1)/2$ constraints and k^2 variables. To solve Thurstone's crime study (Thurstone, 1927b), a relatively large scaling problem involving 19 stimuli, would require solving a linear program with 361 variables and 171 constraints, a significant computational task.

We rewrite the linear program below using matrix notation:

$$\text{Maximize} \quad -(1\underline{u} + 1\underline{v})$$

(14)   subject to $A\underline{s}' + I\underline{u} - I\underline{v} = \underline{x}' \qquad : \underline{\pi}$

$$\underline{u}, \underline{v} \geqslant 0, \quad \underline{s}' \text{ unconstrained}$$

where $\underline{1}$ indicates a row vector of ones, and the objective function has been multiplied by -1 in order to cast the problem as a maximization. Because the origin for the scale estimates is arbitrarily set, it is appropriate to leave the s'_j unconstrained in sign.

We now take the dual of (14), and find that we can exploit the special structure of the constraint matrix, A:

$$\text{Minimize} \quad \underline{\pi} \, \underline{x}'$$

        subject to $\underline{\pi} A = 0 : \underline{s}'$

(15) $\underline{\pi} \geqslant -1 : \underline{u}$

                        $-\underline{\pi} \geqslant -1 : \underline{v}$

The linear program in (15) above is a capacitated network flow problem, as A^T is the appropriate matrix for a complete network with k nodes. The vector $\underline{\pi}$, constrained to the interval $[-1,1]$, is the vector of arc flows. The dual variables $\underline{s}'$ associated with the equality constraints in (15) are the scale estimates.

Thus, it is possible to reduce the T_1 scaling problem for k objects to a capacitated minimum cost network flow problem of k nodes. Thurstone's crime study, mentioned above, would require only 171 variables and 19 constraints, which is not considered a very large network problem. Using the widely

29

accepted network packages, such as GNET (Bradley, Brown, and

Graves, 1975), which exploit the special structure of the network

basis, the problem can be solved quickly and efficiently.

In order to get some idea of how well T_1 scaling does

in recovering the true scale values of a model, we can appeal to

this network conceptualization. Figure III.3 below shows the

contrived five-stimulus example in network form. The sources and

sinks of network flow have been added so that we may refer to arc

flow as a non-negative quantity in the interval $[0,2]$ instead of

$[-1,1]$. The costs on the arcs are the observed values x'_{ij}, and

the reduced costs are denoted $\bar{x}'_{ij}$. For a more detailed

discussion of network flow problems, associated terminology, and

solution procedures, the reader should refer to the relevant

chapters in Bradley, Hax, and Magnanti or Shapiro.

Using this conceptual framework, we can make several

statements about the performance of T_1 scaling:

1. <u>With infallible data, the T_1 procedure estimates the true</u>

<u>scale values exactly for any basic feasible solution to the</u>

<u>network flow problem.</u>

This is seen easily by configuring the network in a straight

line (as shown in Figure III.3 above), positioning the nodes from

left to right in the order indicated by the true scale values for

the stimuli they represent. The result is a network with a cost

on the arc directed from node i to node j equal to the true

difference between the two scale values. Because the dual

variables (which are the scale estimates <u>s'</u>) are determined from

the set of equations $s'_j - s'_i = x'_{ij}$ for all arcs (i,j) in the
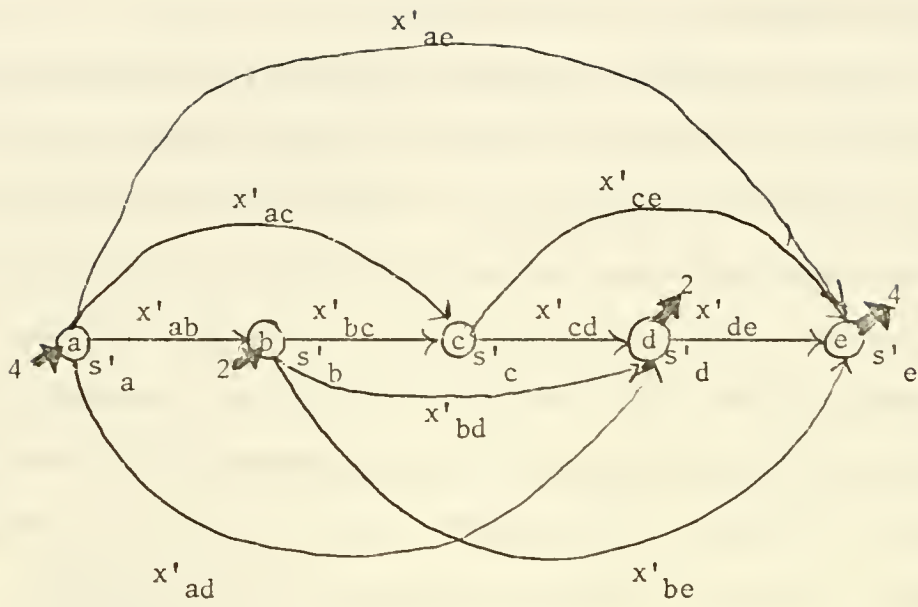
$$s'_a, s'_b, s'_c, s'_d, s'_e \qquad \text{Dual variables (scale estimates)}$$

$$x'_{ab}, x'_{ac}, x'_{ad}, \dots \qquad \text{Arc costs}$$

$$\pi_{ab}, \pi_{ac}, \pi_{ad}, \pi_{ae}, \dots \qquad \text{Arc flows}$$

FIGURE III.3

Network representation of the T_1 scaling problem for the
contrived five-stimulus example of section II.

basis, any spanning tree solution yields the same set of dual variables: always the true scale values, with arbitrary origin.

2. <u>With one inconsistent observation between stimuli i and j in an otherwise infallible set of data, the T_1 procedure always recovers the true scale values.</u>

We demonstrate this result by representing the inconsistent observation as an arc cost between node i and node j equal to the true value x'_{ij} plus some perturbation factor Δ. Without loss of generality, let us choose our initial basic feasible solution so that the arc directed from node i to node j is non-basic and at its lower bound. If Δ is equal to zero, then our data is infallible, and by statement 1 above, any spanning tree solution renders the exact scale estimates. If Δ becomes positive on arc (i,j), then we have no motivation to change our present solution and the dual variables which are the scale estimates remain unaltered. If Δ becomes negative, then we can reduce the cost of our present solution by using the arc at some positive flow capacity.

Once arc (i,j) enters the basis, it reaches full capacity and subsequently leaves the basis. Otherwise, at least one of the dual variables s' will reflect the perturbation factor Δ, and the reduced costs for the non-basic will indicated entry into the basis. When arc (i,j) becomes non-basic at its upper bound, the remaining spanning tree includes only the arcs with perfectly accurate observations. Hence, the dual variables are again the exact scale values, and all reduced costs are zero except for $\bar{x}'_{ij}$, which is negative and at its upper bound.

3. <u>With a small number of inconsistent observations (<<k)</u>
<u>involving mutually independent stimulus pairs in an otherwise</u>
<u>infallible set of data, the T_1 procedure recovers the true scale</u>
<u>values so long as it finds a basic feasible solution where:</u>

    a. <u>all the arcs representing the inaccurate observations</u>
<u>are non-basic.</u>

    b. <u>all of the arcs representing the observations that are</u>
<u>higher than their true value are at their lower bound.</u>

    c. <u>all of the arcs representing the observations that are</u>
<u>lower than their true value are at their upper bound.</u>

      This result follows from the line of analysis pursued
in statement 2 above. Clearly, so long as the method finds such
a spanning tree solution, all reduced costs for the non-basic
arcs representing inaccurate observations higher than their true
values will be positive, and reduced costs for the non-basic arcs
at their upper bound will be negative. The dual variables for
this solution will be determined from the costs of the basic
arcs, which are all accurate observations; thus, the true scale
values will be recovered.

      Once the inaccuracies in the observation begin to
affect stimulus pairs with common elements (such as i,j and i,k),
it is difficult to determine how the error is affecting the
resulting scale estimates. Our diagnostic is unable to isolate
these instances of "overlapping" error, as it functions to adjust
only two scale estimates at a time. Because scaling problems
typically involve a rather small number of items anyway, more
than one or two serious inaccuracies indicates the possibility of

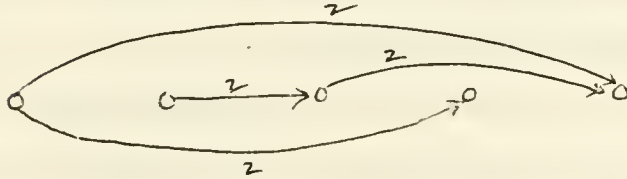some violation of the assumptions of the scaling model.

It remains to be seen, through some sort of empirical validation, just how often the T_1 procedure does find the "uncontaminated" spanning tree solution, and further investigation in this area is indicated.

We applied the T_1 procedure to the contrived five-stimulus example. As anticipated, it recovered the true scale values of the errorless model, except for a translation in scale origin. The solution procedure for the five node minimum cost network flow problem is shown in four steps in Figure III.4. Initially, the arc representing the inconsistent observation between stimuli a and c is non-basic. The associated cost is .15 + Δ , where .15 is the coefficient for the errorless model, and the perturbation factor Δ equals -.25. Step 1 indicates that for $\Delta < 0$, arc (a,c) should enter the basis. Once arc (a,c) has entered the basis, still at its lower bound, several other arcs become candidates to enter the basis, as shown in Step 2. Only when (a,c) leaves the basis in Step 4 do we reach an optimal (once again degenerate) solution. Mosteller's goodness-of-fit test reveals a negligible difference between θ' and θ" for all stimulus pairs except a,c; the T_1 solution confines 100% of the error to the single stimulus pair previously identified by the diagnostic as suspicious, and does so regardless of the magnitude of the error factor Δ .

In summary, the T_1 scaling procedure appears to be a highly desirable alternative to Thurstone's least squares approach. Although computationally more time consuming, the T_1
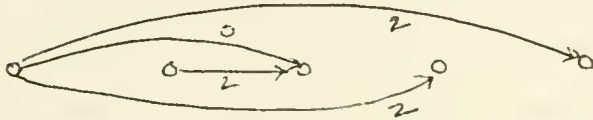
34

STEP 1

$S'_a = 0$
$S'_b = .10$
$S'_c = .15$
$S'_d = -.15$
$S'_e = .20$

$\bar{\bar{x}}'_{ab} = 0 \ (L)$ $\bar{x}'_{ac} = +\Delta$ *
$\bar{x}'_{bd} = 0 \ (L)$ $\bar{x}'_{cd} = 0 \ (L)$
$\bar{x}'_{be} = 0 \ (L)$ $\bar{x}'_{de} = 0 \ (L)$
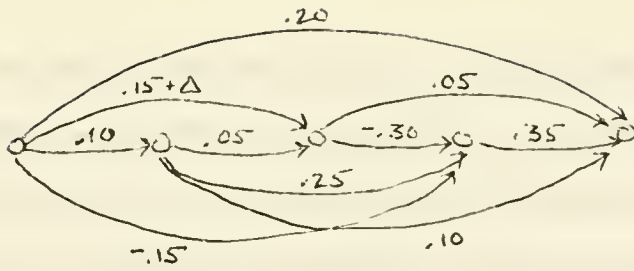
ALL BASIC ARCS AT
UPPER BOUND

* FOR $\Delta < 0$, ARC (a,c)
ENTERS BASIS
ARC (c,e) DROPS OUT
AT UPPER BOUND

STEP 2.

$S'_a = 0$
$S'_b = .1 + \Delta$
$S'_c = .15 + \Delta$
$S'_d = -.15$
$S'_e = .20$

$\bar{x}'_{ab} = -\Delta \ (L)$ $\bar{x}'_{bd} = \Delta$ *
$\bar{x}'_{ce} = \Delta \ (U)$ $\bar{x}'_{be} = \Delta$ *
$\bar{x}'_{de} = 0 \ (L)$ $\bar{x}'_{cd} = \Delta$ *

* ALL THREE STARRED
ARCS ARE CANDIDATES
TO ENTER BASIS,
ADD (b,d); DROP (a,d)
AT LOWER BOUND.

STEP 3

$S'_a = 0$
$S'_b = .1 + \Delta$
$S'_c = .15 + \Delta$
$S'_d = -.15 + \Delta$
$S'_e = .20$
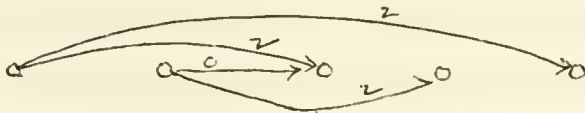
$\bar{x}'_{ab} = -\Delta \ (L)$ $\bar{x}'_{ad} = -\Delta \ (L)$
$\bar{x}'_{ce} = \Delta \ (U)$ $\bar{x}'_{be} = \Delta$ *
$\bar{x}'_{cd} = 0 \ (L)$ $\bar{x}'_{de} = \Delta$ *

* ADD ARC (b,e)
ARC (a,c) REACHES
UPPER BOUND AND
DROPS FROM BASIS

STEP 4.

$S'_a = 0$
$S'_b = .10$
$S'_c = .15$
$S'_d = -.15$
$S'_e = .20$
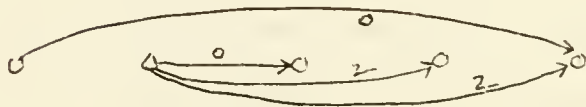
$\bar{x}'_{ab} = 0 \ (L)$ $\bar{x}'_{ac} = \Delta \ (U)$
$\bar{x}'_{ce} = 0 \ (U)$ $\bar{x}'_{cd} = 0 \ (L)$
$\bar{x}'_{ad} = 0 \ (L)$ $\bar{x}'_{de} = 0 \ (L)$

OPTIMAL

FIGURE III.4.   Four steps of the minimum cost network flow solution
procedure for the five-stimulus example.

35

scaling method can use existing network packages to solve very large scaling problems (around 20 objects) in seconds. When the observed data reflect exactly the true form of the model, T_1 scaling and least squares estimate the scale values exactly. When a serious inaccuracy is present in an otherwise accurate set of data, the T_1 procedure can still recover the true scale values, whereas least squares cannot always. It is also important to note that T_1 scaling does not fall prey to the same weakness that L_1 regression does: no matter how inaccurate the one bad observation is, T_1 scaling still recovers the true scale values of the model.

IV. Results from Applying the T_1 Procedure to Scaling Problems from the Literature

We applied the sensitivity diagnostic and the T_1 procedure to certain Thurstonian scaling problems from the literature, in order to determine their collective effectiveness in identifying and resolving potential trouble with real paired comparisons data. The first case involves a subset of the 1948 American League baseball data presented by Mosteller (1951). Each one of five teams -- Cleveland, Boston, New York, Washington, and Chicago -- played 22 games against each of the four others. The proportion of games each team won from the other team, analogous to the proportion of times one team is judged better than another, is shown in Chart IV.1.

The least squares scale estimates from the Thurstonian scaling model (shown in Chart IV.1) reveal a rather disappointing configuration. The scale is split by a wide, empty interval, with Chicago and Washington lumped together at the low end of the scale and Boston, Cleveland, and New York almost on top of one another at the high end.

Kendall and Smith's coefficient of consistency for these data is .80, indicating some element of intransitivity in this collective ordering. The one circular triad in the data occurs with New York, Cleveland, and Boston: New York won over 50% of the games it played against Cleveland, Cleveland won over 50% of its games against Boston, and yet Boston won over 60% of its games against New York. A coefficient value of .80, however, does not conclusively demonstrate the failure of the comparisons

37

method to systematically discriminate between teams; the

probability that one or fewer intransitive triads occur due to an

unsystematic (random) series of judgments is low, less than .24.

If it is not the entire method which is at fault, then the error

might be due to a problematic comparison between two teams that

is interfering with the process of determining scale estimates

for the "best" of the five teams.

The sensitivity diagnostic reveals that almost 80% of

the discrepancy in fit is eliminated if the scale estimates for

New York and Boston are adjusted. One possible reason for this

error is that some exogenous factor operated on the series of

games between these two teams to produce results inconsistent

with the other series. During the era of the Yankees' supremacy

in the American League, it was often said: "The New York Yankees

are the champions of the world, but the Red Sox are champions of

the Yankees," because the Red Sox seemed able to beat the Yankees

fairly consistently, even though the Yankees at that time had the

best all-around record in baseball. On the assumption that this

discrepancy might have been due to home field conditions,

"rivalry", or a variety of other exogenous conditions not common

to the series played between the other teams, we applied the T_1

scaling procedure and assessed the ultimate effect on the fit of

the model.

The T_1 scale estimates shown in Chart IV.1 reflect a

noticeably different configuration. The scale positions for New

York and Boston, which were almost identical in the least squares

solution, are now widely separated, clearly identifying New York

1. Cleveland
2. Boston
3. New York   Identification of Scale Symbols
4. Washington
5. Chicago

1.	x				
2.	.522	x			
3.	.455	.636	x		
4.	.727	.682	.773	x	
5.	.727	.636	.727	.571	x

Matrix of Observed Proportions

5(-.35)   4(-.33)   2(.22)
                                           3(.22) 1(.23)
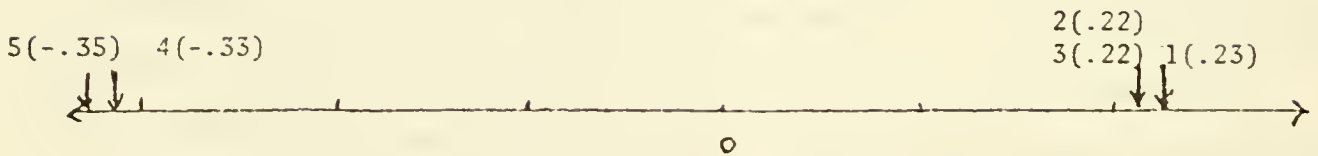
Thurstonian Scale Estimates

1.	x				
2.	1.2	x			
3.	7.4	62.6	x		
4.	0.9	2.9	17.2	x	
5.	0.3	23.7	0.5	13.6	x

Matrix of $(\theta''_{ij} - \theta'_{ij})^2$ from Mosteller's fit criterion

CHART IV.1

1.   x

2.   127.7   x   Remaining discrepancy in fit after

3.   116.9 | 29.2 | x   sensitivity diagnostic applied to each

4.   128.4   124.4   98.6   x   given stimulus pair

5.   128.9   88.4   129.4   108.3   x

T_1 Scale Estimates

1.   x

2.   0.0   x

3.   0.2 | 148.3 | x   Matrix of $(\theta_{ij} - \hat{\theta}_{ij})^2$ from

4.   2.7   0.0   3.4   x   Mosteller's fit criterion.

5.   8.3   2.3   0.0   29.0   x

CHART IV.1 (cont'd)

as the "best" team of the five. These results are quite similar
to the five-stimulus example of section II. The T_1 procedure
produces scale estimates that largely ignore the single
substantial source of inconsistency in the model. In the
baseball data, where one series of games should not have
inordinate influence on the determination of the "best" team, the
T_1 procedure provides a solution that is informative and
conceptually appealing.

The second example, a scaling of attitude statements on
the participation of the United States in the Korean War, is
taken from Hill (1953). Hill selected a subset of seven
prescaled attitude statements that he deliberately biased toward
the favorable side. His hypothesis was that where statements
were conceptually closer on the "favorable/unfavorable"
continuum, there would be more inconsistency reflected in the
observational data. Kendall and Smith's coefficient equals
unity, indicating the absence of any circular triad. Clearly,
however, there is some form of inconsistency affecting the data.
Mosteller's fit criterion for the least squares estimates is poor
for n = 94 comparisons for each stimulus pair.

The sensitivity diagnostic reveals that no single
stimulus pair reduces the discrepancy in fit by more than 25%;
using the diagnostic to adjust any one of 16 out of 21 pairs in
the lower diagonal of X' does not reduce the error by more than
10%. Thus, the sensitivity procedure does not identify any
source of limited intransitivity in the data.

The T_1 scale estimates are largely similar to those

39

given by the least squares approach, indicating that when the source of inconsistency in the data is not limited or well defined, the T_1 procedure does at least as well in estimating the true scale values as the Thurstonian least squares. The results are presented in Chart IV.2 below.

In conclusion, we can use the sensitivity diagnostic presented above to determine where problems with inconsistency appear in the observed data, how much the occurrence of inconsistency degrades the fit of the model, and the nature of the distortion of the scale estimates. This sensitivity analysis involves discounting a single stimulus pair at a time, and the method is straightforward and computationally simple. The results provide a better idea of problems within the data and indicate when there is a need for more robust scale estimates that discount these data problems.

These more robust estimates may be obtained by solving the minimum cost network flow problem outlined above as T_1 scaling. The procedure provides scale estimates that are not inordinately influenced by the presence of limited sources of inaccuracy in the data.

1. I suppose the U.S. has no choice but to continue the Korean War.
2. We should be willing to give our allies in Korea more money if they need it.
3. Withdrawing our troops from Korea at this time would only make matters worse.
4. The Korean War might not be the best war to stop communism, but it was the only thing we could do.
5. Winning the Korean War is absolutely necessary whatever the cost.
6. We are protecting the United States by fighting in Korea.
7. The reason we are in Korea is to defend freedom.

Identification of Scale Symbols

	1	2	3	4	5	6	7	
1.	x							
2.	.309	x						
3.	.202	.457	x					
4.	.149	.426	.479	x				Matrix of Observed Proportions
5.	.202	.340	.372	.479	x			
6.	.085	.277	.362	.330	.457	x		
7.	.064	.138	.330	.287	.415	.394	x	

4(.03)

1(-.87)   2(-.27)   3(-.03) 5(.17)   6(.37)   7(.59)

o   Thurstonian Scale Estimates
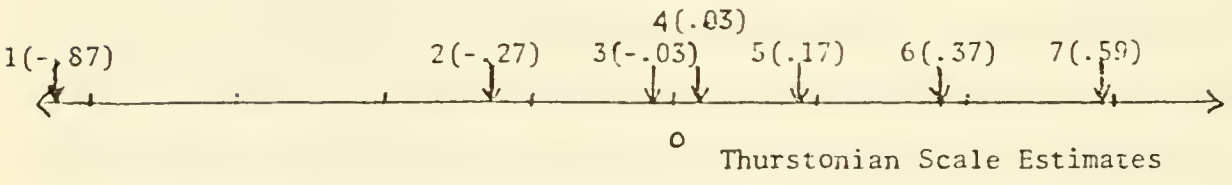
	1	2	3	4	5	6	7	
1.	x							
2.	5.0	x						
3.	0.0	8.6	x					
4.	6.9	6.8	0.1	x				Matrix of $(\theta_{ij} - \theta_{ij})^2$
5.	18.7	0.9	5.9	5.3	x			From Mosteller's fit criterion
6.	4.6	1.1	1.3	5.1	3.0	x		
7.	0.9	20.2	14.9	0.0	17.1	1.5	x	

CHART IV.1

1.	x						
2.	120.6	x					
3.	128.0	117.7	x				
4.	116.3	120.4	127.9	x			
5.	103.0	126.6	121.4	118.4	x		
6.	119.6	127.6	126.5	120.3	125.5	x	
7.	125.2	94.6	109.1	127.9	108.0	125.1	x

Remaining discrepancy in fit after sensitivity diagnostic applied to each given stimulus pair.

$1(-.85)$ $2(-.27)$ $3(-.02)$ $4(.03)$ $5(.16)$ $6(.34)$ $7(.60)$

0

T_1 Scale Estimates

1.	x						
2.	4.4	x					
3.	0.0	8.7	x				
4.	8.7	5.6	0.0	x			
5.	11.5	0.0	11.9	2.4	x		
6.	9.6	0.0	0.0	9.7	2.6	x	
7.	1.2	20.5	14.5	0.0	25.7	0.0	x

Matrix of $(\theta_{ij} - \theta_{ij})^2$ from Mosteller's fit criterion

CHART IV.2 (cont'd)

REFERENCES

Barrodale, J. "L_1 approximation and the analysis of data," Applied
    Statistics, vol. 17 (1968). pp. 51 - 57.

Barrodale, I. and Roberts, F.D.K., "An Improved Algorithm for Discrete
    L_i Linear Approximation," SIAM Journal of Numerical Analysis,
    1973, pp.839 - 848.

Belsley, Kuh, and Welsch, Regression Diagnostics. New York; Wiley, 1980.

Bradley, Brown, and Graves, GNET: General Network Package. Western
    Management Science Institute, California; 1975.

Bradley, Hax, and Magnanti, Applied Mathematic Programming. Mass;
    Addison-Wesley, 1977.

Hill, Richard J. "A Note on Inconsistency in Paired Comparison Judgments,"
    American Sociological Review, 1953, pp. 564 - 566.

Kendall, M.G. and Smith, B. Babington, "On the Method of Paired Comparisons,"
    Biometrika, vol. 31, (1940). pp. 324 - 345.

Krasker, W. S. and Welsch, Roy, "Efficient Bounded-Influence Regression
    Estimation Using Alternative Definitions of Sensitivity," Technical
    Report #TR-3, Center for Computational Research, MIT, December, 1979.

Mosteller, F. "Remarks on the Method of Paired Comparisons: III. A Test
    of Significance for Paired Comparisons when Equal Standard Deviations
    and Equal Correlations are Assumed," Psychometrica, 1951, pp. 207 - 218.

Shapiro, J. Mathematical Programming: Structures and Algorithms.

    New York; Wiley, 1979.

Thurstone, L. L. "A Law of Comparative Judgment," Psychology Review,

    1927, pp. 273 - 286.

_____, "The Method of Paired Comparisons for Social Values,"
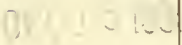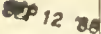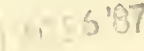
    Journal Of Abnormal and Social Psychology, vol 21 (1927), pp. 384 - 400.

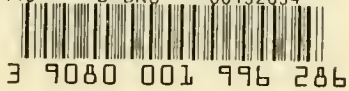Torgerson, W. Theory and Methods of Scaling, New York; Wiley, 1958.