PSYCHOMETRIKA—VOL. 40, NO. 2
JUNE, 1975

A SIMPLE METHOD FOR PAIRWISE MONOTONE REGRESSION

RICHARD M. JOHNSON

MARKET FACTS, INC.

A simple method of monotone regression is described based on the principle of minimizing pairwise departures from monotonicity.

Background

In recent years the practical usefulness of conjoint measurement techniques has been firmly established, particularly in the field of consumer research [Davidson, 1973; Fiedler, 1972; Green and Rao, 1971; Johnson, 1974]. A model of preference formation in common use is the simple additive one in which each possible level of an attribute is considered to have a "part value" to an individual, and where the "total value" to him of an object is the sum of the part values of its attributes. Most frequently the input data for analysis consist of rank orders of preference among objects which differ in known ways on several attributes.

A group of somewhat different models are concerned with relations between an individual's preferences for objects and their locations in a space. These models focus either on the object locations themselves, or on discrepancies between objects' positions and an ideal level for each dimension. Models of this type are discussed by Carroll [1972], Srinivasan and Shocker [1973], and Peckelman and Sen [1974]. In these models the input consists of individuals' rank orders of preference among objects, together with spatial coordinates of the objects.

For models of either type some regression-like procedure is required to estimate either part values, dimensional weights, or ideal point locations. Since the input data are scaled only at the rank order level, standard regression techniques are often inadequate. However, "nonmetric" or "monotone" regression can be used. (The terminology "isotonic regression" is preferred by Barlow, *et al.* [1972]).

In the conjoint measurement case a coefficient matrix of dummy variables, consisting of zeroes and ones, can be used to indicate presence or absence in an object of each level of each attribute. This matrix might have a row for each object and a column for each attribute level. A unit element in the i, jth position would indicate that object i had the jth attri-

bute level. For spatial models the coefficient matrix might contain object scores for each dimension, or squared distances between object positions and known ideal point locations for each dimension. In either case, a set of weights would be sought, one for each column, so that the weighted row sums of the coefficient matrix would be monotonic with the individual's rank order of preference among the objects described by that matrix. The weights might then be interpreted as the part value to that individual of each attribute level in the conjoint measurement case, or the importance of each dimension or attribute in the case of the spatial models.

One computational algorithm for this purpose is Kruskal's MONANOVA [1965]. Although actually a nonmetric analog of analysis of variance, that procedure may be regarded more generally as a special case of monotone regression. MONANOVA attempts to minimize the same badness of fit measure, "stress", employed in Kruskal's well-known nonmetric multidimensional scaling procedure MDSCAL [1964].

Johnson [1973] has described a nonmetric scaling method which, although appearing to produce results substantially similar to Kruskal's, attempts to minimize a badness of fit measure simpler in concept than stress. Just as the stress measure forms a natural foundation for additional nonmetric procedures such as MONANOVA, so the badness of fit measure described by Johnson can be extended easily to form the basis for the nonmetric monotone regression procedure described here.

Quantifying Lack of Fit

Consider the matrix $\mathbf{X}$ of order $n \times p$ containing values for n objects on p independent or perhaps dummy variables. Let the vector $\mathbf{y}$ of length n contain an individual's preference ratings or rankings for the same n objects. Consider an unknown vector $\mathbf{b}$, of length p containing "weights." Let $\mathbf{Xb} = \hat{\mathbf{y}}$. Then the monotone regression problem may be described as that of finding a vector $\mathbf{b}$ so that the elements of the fitted vector $\hat{\mathbf{y}}$ are as nearly monotonic with corresponding elements of the given vector $\mathbf{y}$ as possible.

As a measure of the monotonicity of elements of $\hat{\mathbf{y}}$ with those of $\mathbf{y}$, we may use the measure θ, the square of which is defined by Johnson [1973]:

$$\theta^2 = \frac{\sum_{i,j} \delta_{ij}(\hat{y}_i - \hat{y}_j)^2}{\sum_{i,j} (\hat{y}_i - \hat{y}_j)^2} \tag{1}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if sign } (\hat{y}_i - \hat{y}_j) \neq \text{sign } (y_i - y_j) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The numerator of θ^2 is the sum of squared differences between all pairs of "predicted" values which are "in the wrong order." The denominator of θ^2, the sum of all squared differences, is a normalizing constant which confines θ^2 to the unit interval. It can be shown that θ^2 has a natural interpretation as the proportion of the variation among the $\hat{y}$'s which is "inconsistent" with the y's. The statistic θ^2 will be zero with a perfect fit (in the rank order sense) and would have value of 1.0 if the rank order of the "predictions" were exactly the opposite of the rank order of the input data.

If input values are tied it is possible either to allow corresponding fitted values to differ, or to attempt to force ties among them. This is accomplished by redefining δ_{ij} as follows:

$$\begin{aligned} &\delta_{ij} = 1 \quad \text{if } y_i = y_j \text{ and ties are to be "forced."} \\ &\delta_{ij} = 0 \quad \text{if } y_i = y_j \text{ and ties are not to be forced.} \end{aligned} \tag{3}$$

Computation

A simple iterative procedure for the minimization of θ consists of starting with an arbitrary $\mathbf{b}$ and modifying this vector successively in the directions indicated by corresponding gradient vectors.

The gradient vector $\mathbf{g}$ corresponding to $\mathbf{b}$ for any iteration is easy to compute. It may be derived by differentiating θ^2 partially with respect to $\mathbf{b}'$.

Following Johnson [1973], if we let

$$\theta^2 = \frac{u}{v} \tag{4}$$

then

$$\frac{\partial \theta^2}{\partial \mathbf{b}'} = \frac{1}{v^2}\left[v\frac{\partial u}{\partial \mathbf{b}'} - u\frac{\partial v}{\partial \mathbf{b}'}\right]. \tag{5}$$

Using standard calculus it can be shown that

$$\frac{\partial u}{\partial \mathbf{b}'} = 2\sum_{i,j}\delta_{ij}(\mathbf{x}_i - \mathbf{x}_j)(\hat{y}_i - \hat{y}_j) \tag{6}$$

and

$$\frac{\partial v}{\partial \mathbf{b}'} = 2\sum_{i,j}(\mathbf{x}_i - \mathbf{x}_j)(\hat{y}_i - \hat{y}_j) \tag{7}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the ith and jth rows of $\mathbf{X}$, and $\hat{y}_i$ and $\hat{y}_j$ are scalars.

Substituting from (6) and (7) into (5) and simplifying, we get

$$\mathbf{g} = \frac{\partial \theta^2}{\partial \mathbf{b}'} = \frac{2}{v}\sum_{i,j}(\delta_{ij} - \theta^2)(\mathbf{x}_i - \mathbf{x}_j)(\hat{y}_i - \hat{y}_j). \tag{8}$$

It is useful to normalize both $\mathbf{b}$ and $\mathbf{g}$ to have unit sums of squares

at each stage and to use the current value of θ as a "step size" with the recursive relation

$$\mathbf{b}_{m+1} = \mathbf{b}_m - \theta_m \mathbf{g}_m \tag{9}$$

where m indicates iteration number. The process can be terminated when θ stabilizes or after a limiting number of iterations. This simple procedure has appeared to be adequate in those cases examined by the author, although a more sophisticated determination of step size would doubtless provide more rapid convergence.

An Example

The pairwise approach was compared with MONANOVA using a small synthetic example. A 5 × 5 matrix **C** was first constructed with $c_{ij} = 0.11\, i + 0.10\, j + 0.10$. Using a two-way ANOVA model these data can be accounted for perfectly by row and column main effects. Next a second matrix, **D**, was constructed by applying the exponential transformation, $d_{ij} = e^{c_{ij}}$.

Since the elements of **D** are monotonic with those of **C**, both MONANOVA and the pairwise method should produce a perfect fit in a rank order sense, using a row and column main effects model. Such was indeed the case. The coefficient matrix **X** consisted of a 25 by 10 table of zeroes and ones, with rows corresponding to elements of **D** and columns corresponding to main effects to be estimated. Each row of **X** was zero except for two unit elements in the columns corresponding to the main effects affecting that element of **D**.

Next, the elements of **D** were perturbed by adding random numbers rectangularly distributed on the unit interval. The amount of error thus introduced was sizeable, the median perturbation being about 25%. Elements of the resulting matrix were finally rank ordered and submitted to both computing procedures.

The input rank orders are given in Table 1. MONANOVA was able to achieve a final stress of 39.1 percent, and the pairwise method yielded a θ value of .202. The row and column effects estimated by each technique are presented in Table 2, after linear transformations which scaled them so as to most closely approximate the "true" main effects in a least squares sense.

The two methods produced solutions which, though not identical, were approximately equally good in estimating the "true" main effects. Root mean square error values are shown for each solution, and these are approximately equal.

A brief FORTRAN IV program, consisting of fewer than 200 statements, is available from the author. Computing times tend to be approximately linear in p and quadratic in n. For $n = 36$ and $p = 10$ the computing time is approximately 10 iterations per second on a Univac 1108.

Table 1

Input Matrix of Ranks

3	1	6	13	10
2	15	9	7	23
16	4	12	20	19
5	18	17	11	24
8	21	14	22	25

Table 2

Estimated Main Effects

Row	True	MONANOVA	Pairwise
1	-.22	-.188	-.191
2	-.11	-.029	-.014
3	.00	.025	-.014
4	.11	.002	.080
5	.22	.190	.138
Column			
1	-.20	-.188	-.235
2	-.10	-.081	-.058
3	.00	-.066	-.062
4	.10	.077	.120
5	.20	.258	.236
rms error		.054	.051

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. Statistical inference under order restrictions. Wiley, 1972.

Carroll, J. D. Individual differences and multidimensional scaling. In Romney, A., Shepard, R., and Nerlove, S. (eds.), *Multidimensional Scaling: Theory and applications in the behavioral sciences* (Vol. 1, Theory). Seminar Press, 1972.

Davidson, J. D. Forecasting traffic on STOL. *Operational Research Quarterly.* 1973, **24**, 561–569.

Fiedler, J. A. Condominium design and pricing: a case study in consumer trade-off analysis. *Proceedings.* Third annual conference of the Association for Consumer Research, 1972, 279–293.

Green, P. E., and Rao, V. Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research,* 1971, **8**, 355–363.

Johnson, R. M. Pairwise nonmetric multidimensional scaling. *Psychometrika,* 1973, **38**, 11–18.

Johnson, R. M. Trade-off analysis of consumer values. *Journal of Marketing Research,* 1974, **11**, 121–127.

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

Kruskal, J. B. Nonmetric multidimensional scaling: a numerical method. *Psychometrika,* 1964, **29**, 115–129 (b).

Kruskal, J. B. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, Series B,* 1965, **27**, 251–263.

Peckelman, D. and Sen, S. Mathematical programming models for the determination of attribute weights. *Management Science,* 1974, **20**, 1217–1229.

Srinivasan, J. and Shocker, A. D. Linear programming techniques for multidimensional analysis of preferences. *Psychometrika,* 1973, **38**, 337–370.

Manuscript received 9/4/73
Revised manuscript received 12/6/74