PSYCHOMETRIKA—VOL. 54, NO. 3, 371–384
SEPTEMBER 1989

MODEL MODIFICATION

DAG SÖRBOM

UNIVERSITY OF UPPSALA

An analysis of empirical data often leads to a rejection of a hypothesized model, even if the researcher has spent considerable efforts in including all available information in the formulation of the model. Thus, the researcher must reformulate the model in some way, but in most instances there is, at least theoretically, an overwhelming number of possible actions that could be taken. In this paper a "modification index" will be discussed which should serve as a guide in the search for a "better" model. In statistical terms, the index measures how much we will be able to reduce the discrepancy between model and data, as defined by a general fit function, when one parameter is added or freed or when one equality constraint is relaxed. The modification index discussed in this paper is an improvement of the one incorporated in the LISREL V computer program in that it takes into account changes in all the parameters of the model when one particular parameter is freed.

Key words: model modification, maximum likelihood estimation, factor analysis, structural equation models, step-wise regression.

Introduction

Most model fitting techniques take the following general form: Minimize (or maximize) some particular fit function $f(\boldsymbol{\theta}, \boldsymbol{\theta}_c, \mathbf{x})$ with respect to the free parameters $\boldsymbol{\theta}$ and a set of constrained parameters, $\boldsymbol{\theta}_c$, for a given set of data, $\mathbf{x}$. If the model does not fit, as judged by specified criteria, the question arises: How should the model be modified to improve the fit? In this paper we shall consider this question in the case where each parameter in $\boldsymbol{\theta}_c$ is either fixed a priori or specified to be equal to one of the parameters in $\boldsymbol{\theta}$.

Consider an example. In confirmatory factor analysis (Jöreskog, 1969), the parameters in $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_c$ may be taken to be all the elements in the matrix of the factor loadings, $\boldsymbol{\Lambda}$, the covariance matrix of the factors, $\boldsymbol{\Phi}$, and the covariance matrix for the specific factors, $\boldsymbol{\Psi}$. The model specifies that the values of some of these parameters are known a priori (e.g., some $\lambda_{ij} = 0$, $\phi_{kk} = 1$ or, all $\psi_{ij} = 0$, $i \neq j$) or equal to some other elements in these matrices (equality constraints). For maximum likelihood the fit function is

$$f = 2n(\log |\boldsymbol{\Sigma}| + \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - \log |\mathbf{S}| - p), \tag{1}$$

where $\mathbf{S}$ is the sample covariance matrix (a function of the data $\mathbf{x}$), n is the sample size minus one, p is the number of variables and

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}, \tag{2}$$

the covariance matrix for the data implied by the model. If the model does not fit, one may want to know which fixed parameter or which equality constraint should be relaxed to obtain the maximum possible decrease in the value of the function f.

The research reported in this paper has been supported by The Swedish Council for Research in the Humanities and Social Sciences under Research Program "Multivariate Statistical Analysis", Project Director Karl G Jöreskog.

Requests for reprints should be sent to Dag Sörbom, University of Uppsala, Department of Statistics, PO Box 513, S-751 20 Uppsala, SWEDEN.

0033-3123/89/0900-9040$00.75/0

In the general formulation, relaxing a fixed parameter or an equality constraint simply means moving one parameter from $\boldsymbol{\theta}_c$ to $\boldsymbol{\theta}$. The question is: Which one should we choose to maximise the reduction in f? For each parameter in $\boldsymbol{\theta}_c$ we shall construct a *modification index* (MI) which is approximately equal to the decrease in f that will be obtained if this parameter is relaxed. In principle this will be similar to re-estimating each new model, but we will do it by using the currently available information to compute the new estimates explicitly, without having to start the estimating process over and over again.

The problem of model modification has been discussed by, among others, Costner and Schoenberg (1972), Sörbom (1975), Saris, de Pijer, and Zegwaart (1979), Herting and Costner (1985), and MacCallum (1986). The major difficulties with the procedure of Sörbom, which is based on the first order derivatives, are: (i) it does not take into account the different metrics in different parameter matrices (see e.g., Dijkstra, 1981, p. 32), and (ii) it depends on the measurement scales of the observed variables. A procedure which also involves the second-order derivatives was implemented in the computer program LISREL V (see Jöreskog & Sörbom, 1981, p. I.42). That procedure gives a modification index which is a lower bound of the decrease in the fit function. The reason why it gives a lower bound rather than a direct estimate of the decrease is that it evaluates the decrease of the fit function by relaxing one parameter while keeping all the others fixed at their estimated values.

The procedure described in this paper is implemented in the LISREL VI computer program (Jöreskog & Sörbom, 1984). It is a further development in that it will also take into account the change induced in all other parameters if one parameter is relaxed. This means that the procedure will be able to indicate explicitly which parameter constraint should be relaxed in order to maximise the increase in fit. The procedure is equivalent to the score test (Rao, 1948) and the test that results if the Lagrange multiplier theory is applied (see e.g., Aitchison & Silvey, 1958).

In the following section a formal derivation of the modification index will be given. In the subsequent sections the behavior of the index will be illustrated by a number of examples. It should be clear that these illustrations are more or less formal and the indices are used "blindly". In an analysis of real data it is not recommended to use the procedure in this way. It has been done only in order to illustrate the usage and to show how the procedure behaves in ill-conditioned situations, that is, in situations when the information of data is scarce.

Construction of the Modification Index

Let $g = \partial f/\partial\boldsymbol{\theta}$ denote the gradient vector of a fitting function f and let $\mathbf{E} = E[\partial^2 f/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}']$ be the matrix of expected second order derivatives of f. The estimates of $\boldsymbol{\theta}$, denoted $\hat{\boldsymbol{\theta}}$, are those values that minimize $f(\boldsymbol{\theta})$. Let $\mathbf{g}$ and $\mathbf{E}$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ be denoted $\hat{\mathbf{g}}$ and $\hat{\mathbf{E}}$, respectively. Hence we know that $\hat{g} = \mathbf{0}$ and that $\hat{\mathbf{E}}$ can be assumed to be positive definite if the estimated model is identified.

Most fit functions are approximately quadratic around the solution $\hat{\boldsymbol{\theta}}$ and we know that for a sufficiently large sample the f matrix is an approximation of the second order derivatives of f. Thus, we can approximate f around $\hat{\boldsymbol{\theta}}$ by the Taylor expansion

$$f \approx \hat{f} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\hat{g} + \tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\hat{f}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \tag{3}$$

where $\hat{f}$ denotes the value of f at the solution, $\hat{\boldsymbol{\theta}}$. Suppose now that we want to add a previously fixed parameter to the vector of free parameters. Let this new free parameter be denoted by θ_1, and the value to which it was fixed by $\hat{\theta}_1$. The Taylor expansion is

$$f \approx \hat{f} + \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \theta_1 - \hat{\theta}_1 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{g}} \\ \hat{g}_1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \theta_1 - \hat{\theta}_1 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{d}} \\ \hat{\mathbf{d}}' & \hat{k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\theta}}' - \hat{\theta}_1 \end{bmatrix}, \tag{4}$$

where $\hat{g}_1 = \partial f / \partial \theta_1$, $\hat{\mathbf{d}}$ is the vector of expected second order derivatives $E[\partial^2 f / (\partial \boldsymbol{\theta} \partial \theta_1)]$, and $\hat{k}$ is $E[\partial^2 f / \partial \theta_1 \partial \theta_1]$, all first and second order derivatives evaluated at $(\hat{\boldsymbol{\theta}}, \hat{\theta}_1)$.

In order to study how much the function in (4) will be decreased we find new estimates by minimizing f in (4) with respect to $(\boldsymbol{\theta}, \theta_1)$. The minimum, $\hat{\hat{f}}$, must satisfy

$$\begin{bmatrix} \dfrac{\partial f}{\partial \boldsymbol{\theta}} \\ \dfrac{\partial f}{\partial \theta_1} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \tag{5}$$

which, since $\hat{\mathbf{g}} = 0$, becomes

$$\begin{bmatrix} \mathbf{0} \\ \hat{g}_1 \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{d}} \\ \hat{\mathbf{d}}' & \hat{k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \theta_1 - \hat{\theta}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}. \tag{6}$$

Substituting the solution of (6) into (4), we get

$$\hat{\hat{f}} \approx \hat{f} - \begin{bmatrix} \mathbf{0} \\ \hat{g}_1 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{d}} \\ \hat{\mathbf{d}}' & \hat{k} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \hat{g}_1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \mathbf{0} \\ \hat{g}_1 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{d}} \\ \hat{\mathbf{d}}' & \hat{k} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \hat{g}_1 \end{bmatrix} = \hat{f} - \frac{\frac{1}{2}\hat{g}_1^2}{\hat{k} - \hat{\mathbf{d}}'\hat{\mathbf{E}}^{-1}\hat{\mathbf{d}}}. \tag{7}$$

The last step follows by using the formula for the inverse of a partitioned symmetric matrix,

$$\begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{d}} \\ \hat{\mathbf{d}}' & \hat{k} \end{bmatrix}^{-1} = \begin{bmatrix} \hat{\mathbf{E}}^{-1} + m(\hat{\mathbf{E}}^{-1}\hat{\mathbf{d}})(\hat{\mathbf{E}}^{-1}\hat{\mathbf{d}})' & -m\hat{\mathbf{E}}^{-1}\hat{\mathbf{d}} \\ -m(\hat{\mathbf{E}}^{-1}\hat{\mathbf{d}})' & m \end{bmatrix},$$

where $m = 1/(\hat{k} - \hat{\mathbf{d}}'\hat{\mathbf{E}}^{-1}\hat{\mathbf{d}})$. It has been assumed that also the matrix

$$\begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{d}} \\ \hat{\mathbf{d}}' & k \end{bmatrix}$$

is positive definite. This is expected to be the case if the model with θ_1 added as a free parameter is identified. The decrease in function value is $\hat{f} - \hat{\hat{f}}$. Thus, by defining the modification index as

$$\mathrm{MI} = \frac{\frac{1}{2}\hat{g}_1^2}{(\hat{k} - \hat{\mathbf{d}}'\hat{\mathbf{E}}^{-1}\hat{\mathbf{d}})}, \tag{8}$$

we have an approximate estimate of how much the fit function will decrease if one adds a parameter θ_1 previously fixed at $\hat{\theta}_1$ to the set of free parameters. The term $\hat{\mathbf{d}}'\hat{\mathbf{E}}^{-1}\hat{\mathbf{d}}$ in (8) can be seen as an adjustment to the MI that is caused by the change in the parameters in $\boldsymbol{\theta}$ when θ_1 is added. Since $\mathbf{E}$ is positive definite this term is always greater than zero, and this means that we get an increase in the MI when change in $\boldsymbol{\theta}$ is taken into account. In the procedure implemented in the LISREL V computer program this term is not present.

So far we have considered the case of freeing a fixed parameter. We now consider the case when two or more parameters are constrained to be equal and we want to study the effect of relaxing such constraints. Let $\boldsymbol{\theta}_c$ now denote a vector of q parameters all of which are constrained to be equal to the parameter θ_2. This is to say that $\boldsymbol{\theta}_c = \theta_2\mathbf{i}$, where $\mathbf{i}$ is a q-component vector of all ones. In addition, we have, as before, a set of free parameters, $\boldsymbol{\theta}$. We want to compute an estimate of how much the fit function will decrease if we let θ_2 be free while retaining the equality constraints for the elements in $\boldsymbol{\theta}_c$. Suppose f has been minimized with respect to $\boldsymbol{\theta}$ and θ_2 yielding $\hat{\boldsymbol{\theta}}$ and $\hat{\theta}_2$ so that $\hat{\boldsymbol{\theta}}_c = \hat{\theta}_2\mathbf{i}$. The analogy of (4) is

$$f \approx \hat{f} + \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \boldsymbol{\theta}_c - \hat{\boldsymbol{\theta}}_c \\ \theta_2 - \hat{\theta}_2 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{g}} \\ \hat{\mathbf{g}}_c \\ \hat{g}_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \boldsymbol{\theta}_c - \hat{\boldsymbol{\theta}}_c \\ \theta_2 - \hat{\theta}_2 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{D}}_1 & \hat{\mathbf{d}}_2 \\ \hat{\mathbf{D}}_1' & \hat{\mathbf{C}}_{11} & \hat{\mathbf{C}}_{12} \\ \hat{\mathbf{d}}_2' & \hat{\mathbf{C}}_{21} & \hat{C}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \boldsymbol{\theta}_c - \hat{\boldsymbol{\theta}}_c \\ \theta_2 - \hat{\theta}_2 \end{bmatrix}, \tag{9}$$

where $\hat{\mathbf{g}}$, $\hat{\mathbf{g}}_c$ and $\hat{g}_2$ are first order derivatives of f with respect to $\boldsymbol{\theta}$, $\boldsymbol{\theta}_c$, and θ_2. $\mathbf{D}_1$ is the matrix $E[\partial^2 f/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}_c']$, $\hat{\mathbf{d}}_2$ is the vector $E[\partial^2 f/\partial\boldsymbol{\theta}\partial\theta_2]$, $\hat{\mathbf{E}}$ is the expected second order derivatives for the elements in $\boldsymbol{\theta}$ and $\mathbf{C}$ is the matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} E\left[\dfrac{\partial^2 f}{\partial\boldsymbol{\theta}_c\partial\boldsymbol{\theta}_c'}\right] & E\left[\dfrac{\partial^2 f}{\partial\boldsymbol{\theta}_c\partial\theta_2}\right] \\ E\left[\dfrac{\partial^2 f}{\partial\theta_2\partial\boldsymbol{\theta}_c'}\right] & E\left[\dfrac{\partial^2 f}{\partial\theta_2\partial\theta_2}\right] \end{bmatrix}. \tag{10}$$

Substituting $\hat{\boldsymbol{\theta}}_c$ with $\hat{\theta}_2\mathbf{i}$ in (9) and denoting the common element in $\boldsymbol{\theta}_c$ by θ_c we get

$$f \approx \hat{f} + \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \mathbf{i}(\theta_c - \hat{\theta}_2) \\ \theta_2 - \hat{\theta}_2 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{g}} \\ \hat{\mathbf{g}}_c \\ \hat{g}_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \mathbf{i}(\theta_c - \hat{\theta}_2) \\ \theta_2 - \hat{\theta}_2 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{D}}_1 & \hat{\mathbf{d}}_2 \\ \hat{\mathbf{D}}_1' & \hat{\mathbf{C}}_{11} & \hat{\mathbf{C}}_{12} \\ \hat{\mathbf{d}}_2' & \hat{\mathbf{C}}_{21} & \hat{C}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \quad \hat{\boldsymbol{\theta}} \\ \mathbf{i}(\theta_c - \hat{\theta}_2) \\ \theta_2 - \hat{\theta}_2 \end{bmatrix}.$$

By the constraint $\hat{\boldsymbol{\theta}}_c = \hat{\theta}_2\mathbf{i}$ we know that $\mathbf{i}'\hat{\mathbf{g}}_c + \hat{g}_2 = 0$. Thus, by replacing $\mathbf{i}'\hat{\mathbf{g}}_c$ with $-\hat{g}_2$ in (9) and utilizing the fact that $\mathbf{g} = \mathbf{0}$ we get

$$f \approx \hat{f} + \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \theta_c - \hat{\theta}_2 \\ \theta_2 - \hat{\theta}_2 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{0}} \\ -\hat{g}_2 \\ \hat{g}_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \theta_c - \hat{\theta}_2 \\ \theta_2 - \hat{\theta}_2 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{D}}_1\mathbf{i} & \hat{\mathbf{d}}_2 \\ \mathbf{i}'\mathbf{D}_1' & \mathbf{i}'\hat{\mathbf{C}}_{11}\mathbf{i} & \mathbf{i}'\hat{\mathbf{C}}_{12} \\ \hat{\mathbf{d}}_2' & \hat{\mathbf{C}}_{21}\mathbf{i} & \hat{C}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \theta_c - \hat{\theta}_2 \\ \theta_2 - \hat{\theta}_2 \end{bmatrix}.$$

Minimizing f with respect to $\boldsymbol{\theta}$, θ_c, and θ_2, results in an expression for the MI as

$$\text{MI} = \tfrac{1}{2}\hat{g}_2^2 \begin{bmatrix} \hat{\mathbf{0}} \\ -1 \\ 1 \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{E}} & \hat{\mathbf{D}}_1\mathbf{i} & \hat{\mathbf{d}}_2 \\ \mathbf{i}'\hat{\mathbf{D}}_1' & \mathbf{i}'\hat{\mathbf{C}}_{11}\mathbf{i} & \mathbf{i}'\hat{\mathbf{C}}_{12} \\ \mathbf{d}_2' & \hat{\mathbf{C}}_{21}\mathbf{i} & \hat{C}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{0}} \\ -1 \\ 1 \end{bmatrix}. \tag{11}$$

By use of the formula for inversion of a partitioned matrix (11) can be written

$$\text{MI} = \tfrac{1}{2}\hat{g}_2^2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}' \left[\begin{bmatrix} \mathbf{i}'\mathbf{C}_{11}\mathbf{i} & \mathbf{i}'\hat{\mathbf{C}}_{12} \\ \hat{\mathbf{C}}_{21}\mathbf{i} & \hat{C}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{i}'\hat{\mathbf{D}}_1' \\ \hat{\mathbf{d}}_2' \end{bmatrix} \hat{\mathbf{E}}^{-1}(\hat{\mathbf{D}}_1\mathbf{i} \;\; \hat{\mathbf{d}}_2) \right]^{-1} \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \tag{12}$$

Thus, denoting the elements of the matrix to be inverted in (12) by h_{ij} we can write the MI as

$$\text{MI} = \frac{\tfrac{1}{2}\hat{g}_2^2(h_{11} + 2h_{21} + h_{22})}{h_{11}h_{22} - h_{12}^2}. \tag{13}$$

The inverse of $\hat{\mathbf{E}}$ in (12) can be computed from the estimate of the inverse of the expected second order derivatives at the constrained solution. Let the second order derivatives at the solution be denoted

$$\hat{\mathbf{E}}^* = \begin{bmatrix} \mathbf{E} & \mathbf{e}_{12} \\ \mathbf{e}_{21} & \mathbf{e}_{22} \end{bmatrix},$$

where $\mathbf{E}$, as before, is the matrix of expected second order derivatives for the free parameters and

$$\mathbf{e}_{12} = \mathbf{e}_{21}' = \hat{\mathbf{D}}_1\mathbf{i} + \hat{\mathbf{d}}_2$$

$$\mathbf{e}_{22} = \mathbf{i}'\mathbf{C}_{11}\mathbf{i} + \mathbf{i}'\mathbf{C}_{12} + \mathbf{C}_{21}\mathbf{i} + \mathbf{c}_{22}.$$

Thus, by using the formula for the inverse of a partitioned matrix we can get

$$\hat{\mathbf{E}}^{-1} = \mathbf{E}^{11} - \frac{(\mathbf{e}^{12}\mathbf{e}^{21})}{\mathbf{e}^{22}},$$

where the superscripts denote the matrices in the inverse of the partitioned matrix $\mathbf{E}^*$.

In the following examples the fit function associated with the maximum likelihood method will be considered, that is the fitting function is given by (1). If the model is correct and the observed variables have a multivariate normal distribution, then f in (1) is approximately distributed in large samples as central χ^2 with degrees of freedom equal to $p(p + 1)/2 - t$, where t is the number of free parameters of the model. The MI's will then be approximately distributed as χ^2 with 1 degree of freedom, which also holds when the GLS fit function is used.

The assumptions made concerning the distribution of f reveal that there are two major requirements:

1. The initial model has a "reasonable" fit.
2. The sample size is "sufficiently" large.

For a discussion of these requirements and the impact of violations of them, see MacCallum (1986).

The procedure described above was implemented in the LISREL VI computer program (Jöreskog & Sörbom, 1984). This program computes the estimates of the parameters of a model by an iterative minimization of the fit function f. In order to compute standard errors of the parameter estimates, the information matrix, that is, the matrix $\mathbf{E}$ in (3), and its inverse are evaluated at the minimum. Thus, to compute the MI's, we need only compute the vector $\mathbf{d}$ and the scalar k in (8) for each fixed parameter of the model. In addition, we will get estimates of the parameters of the modified model by the solution of (6). This has been subsumed in the computer program by a procedure for automatic modification of a model. The program finds the parameter with the largest MI, adds this to the set of free parameters and goes on repeatedly until none of the MI's is greater than a specified value. Since the MI's are approximately χ^2 with 1 degree of freedom, the value could be chosen such that the procedure stops when there is no MI significant at, for example, the 1%-level. Or, in other words, the procedure stops when no significantly better model can be found. At each step the estimates from the solution of (6) in the previous step are used as starting values for the iterative minimization of f. This procedure has been found to be very efficient, at least numerically. How it works from a substantial point of view must however be judged from case to case.

For a fixed parameter which would be nonidentified if set free, the first order derivative must be zero. If this was not the case, it would mean that we could estimate

TABLE 1.

Covariance matrix for Werner Blood Chemistry Data (N=180).

y	1827.015							
x_1	154.514	97.978						
x_2	1.220	2.192	6.161					
x_3	128.106	51.804	24.093	420.242				
x_4	1.965	0.279	0.204	0.823	0.251			
x_5	0.882	-0.280	-0.005	-1.725	-0.042	0.129		
x_6	5.149	-0.040	0.168	0.627	-0.015	0.077	0.224	
x_7	13.130	2.314	0.349	6.977	0.009	0.012	0.088	1.257

y = Cholesterol x_1 = Age
x_2 = Height x_3 = Weight
x_4 = Birthpill x_5 = Albumin
x_6 = Calcium x_7 = Uric Acid

the model with the parameter free and obtain a decrease in the fit function, since the matrix of expected second order derivatives is positive definite. This contradicts the non-identification status of the parameter, since changing a non-identified parameter can not change the value of the fit function. Thus, by (8) the MI is zero for each such parameter. This in turn means that, as long as we start with an identified model, the procedure for automatic modification should produce identified models.

Stepwise Regression

The MI-procedure can be used to perform forward stepwise regressions as will be illustrated by using data from BMDP (1977). The covariance matrix of the Werner Blood Chemistry data is given in Table 1. The regression model is

$$\mathbf{y} = \boldsymbol{\gamma}'\mathbf{x} + \zeta, \tag{14}$$

where $\mathbf{y}$ is the dependent variable, $\mathbf{x}$ a set of explanatory variables, $\boldsymbol{\gamma}$ a vector of regression coefficients, and ζ the regression residual. If we initially let $\boldsymbol{\gamma}$ be a fixed vector of zeros, we can add explanatory variables one after another in accordance with the size of MI's of the γ's. The maximum MI's at each step are listed in Table 2.

At Step 0 we analyze the model in (14) with the constraint that all γ's are fixed 0's. We find that γ_1 has the largest MI, and hence we include the variable x_1 as an explanatory variable and go on to Step 1, where we find that x_6 should be included and so on. We stop the fitting process as soon as we get a nonsignificant decrease in χ^2 from one

TABLE 2.

Steps taken to modify the model for Blood Chemistry Data (ML)

Step	Parameter to free	Mod. index	Chi-square	d.f.	Prob. level
0	γ_1	23.87	49.58	7	0.000
1	γ_6	13.74	23.06	6	0.001
2	γ_7	5.75	9.66	5	0.085
3	γ_4	1.68	3.81	4	0.432
4	γ_2	1.85	2.13	3	0.547
5	γ_3	0.27	0.27	2	0.875
6	γ_5	0.00	0.00	1	0.970

TABLE 3.

Steps taken to modify the model for Blood Chemistry Data (GLS)

Step	Parameter to free	Mod. index	Chi-square	d. f.	Prob. level
0	γ_7	43.33	61.73	7	0.000
1	γ_1	8.13	18.40	6	0.005
2	γ_6	8.57	10.27	5	0.068
3	γ_4	0.68	1.70	4	0.790
4	γ_2	0.96	1.02	3	0.795
5	γ_3	0.07	0.07	2	0.967
6	γ_5	0.00	0.00	1	0.988

step to the next. Thus, we should stop after step 3, since the decrease in χ^2 is only 1.68, which is not significant. We obtain a regression model with variables x_1, x_6, and x_7, which is in concordance with the result that is obtained if an ordinary forward stepwise regression analysis is performed with, for example, the program BMDP2R (BMDP, 1977).

An examination of Table 2 reveals that the MI is not exactly equal to the decrease in χ^2, but that it is a very good approximation. This is due to the approximation in (3) of the fit function (1). If, instead, we use the GLS fit function

$$f = \mathrm{tr}\,[(\mathbf{S} - \mathbf{\Sigma})\mathbf{S}^{-1})]^2,$$

and treat the **x**-variables as fixed, it can be shown that the approximation in (4) is exact. At each step the only nonzero elements of $\mathbf{S} - \mathbf{\Sigma}$ are $s_{1j} - \sigma_{1j}, j = 2, 3, \ldots, 8$, and, thus, f is a quadratic function of the elements of $\boldsymbol{\gamma}$. This is confirmed in Table 3, where the results from a stepwise regression using the GLS fit function are shown. Here the MI's are exactly equal to the differences in chi-square.

Factor Analysis Examples

As a first factor analysis example, the covariance matrix in Table 4 will be analyzed. This is taken from Costner and Schoenberg (1972) and is a population covariance matrix for the model depicted in Figure 1. The model for the observed x-variables can be written as

$$\mathbf{x} = \mathbf{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{15}$$

TABLE 4.

Population Covariance matrix for the model in Figure 1 (N = 1000).

x_1	1.000						
x_2	0.150	1.000					
x_3	0.210	0.350	1.000				
x_4	0.036	0.060	0.084	1.000			
x_5	0.072	0.120	0.168	0.580	1.000		
x_6	0.108	0.180	0.252	0.520	0.540	1.000	
x_7	0.144	0.240	0.336	0.160	0.320	0.480	1.000

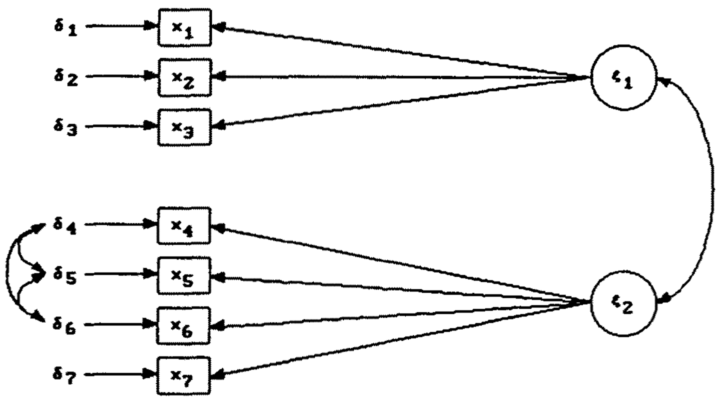

FIGURE 1
Model for the population matrix in Table 4.

where, $\boldsymbol{\xi}$ is a vector of common factors and $\boldsymbol{\delta}$ a vector of unique factors, uncorrelated with $\boldsymbol{\xi}$. The model in (15) implies that the covariance matrix for the x-variables is the $\boldsymbol{\Sigma}$-matrix as given in (2).

For the model in Figure 1 the elements (5,4), (6,4), and (6,5) of $\boldsymbol{\Psi}$ are equal to 0.5, 0.4, and 0.3, respectively. Starting with a model with $\boldsymbol{\Psi}$ diagonal, that is, a model with uncorrelated errors gets a χ^2 equal to 221.8 with 13 degrees of freedom, indicating there is something wrong with the fit of the model. As convincingly shown by Costner and Schoenberg (1972) a serial examination of the residuals, $\boldsymbol{\Sigma} - \mathbf{S}$, may be very misleading. They suggest a 7-step procedure to detect misspecification. We proceed as shown in Table 5A, and arrive at a model with perfect fit, although it is not the one in Figure 1. This is due to the fact that any of the covariances among δ_4, δ_5 and δ_6 in the population model can be set to zero and accounted for by the path from the first factor to variable $\mathbf{x}_7$. Therefore, there are at least four models which are equivalent in the sense that they produce exactly the same covariance matrix, and, hence, the same chi-square. Thus, there is no possibility for a statistical method based on a fit function to discriminate between these models. The procedure simply needs more information in order to find the "correct" model. One way of doing this would be to neglect any suggestion by the procedure to free an element in $\boldsymbol{\Lambda}$. In the LISREL VI program there is an option to specify any fixed parameter to be kept fixed during the search for a model. If this option is used we arrive at the population model after 3 steps, as illustrated in Table 5B.

The MI can be used to illustrate the problem with equivalent models at least in the case when there is just one single misspecification. If, for example, we fix one of the covariances among δ_4, δ_5, and δ_6 in the population model and compute the MI's, we will find that we get exactly the same modification index for the fixed covariance and the factor loading of $\mathbf{x}_7$ and ξ_1.

As a second example, consider the selection of nine mental ability tests from the

TABLE 5A.

Steps taken to modify the model for Population Covariance Matrix Data

Step	Parameter to free	Mod. index	Chi-square	d. f.	Prob. level
0	Ψ_{54}	140.44	221.78	13	0.000
1	λ_{71}	73.47	112.96	12	0.000
2	Ψ_{64}	37.45	38.14	11	0.000
3			0.00	10	1.000

TABLE 5B.

Steps taken to modify the model for Population Covariance Matrix Data

Step	Parameter to free	Mod. index	Chi-square	d. f.	Prob. level
0	Ψ_{54}	140.44	221.78	13	0.000
1	Ψ_{64}	66.38	112.96	12	0.000
2	Ψ_{65}	55.97	54.98	11	0.000
3			0.00	10	1.000

study of Holzinger and Swineford (1939) as given in Table 6. This subset was analyzed by Jöreskog (1969) and it was supposed that the correlations could be accounted for by three factors, namely visualization (Tests 1, 2, and 3), verbal intelligence (Tests 4, 5, and 6), and speed (Tests 7, 8, and 9). Assuming a structure defined by the factor analysis model in (15), the factor loading matrix should have the following simple pattern:

$$\begin{bmatrix} x & 0 & 0 \\ x & 0 & 0 \\ x & 0 & 0 \\ 0 & x & 0 \\ 0 & x & 0 \\ 0 & x & 0 \\ 0 & 0 & x \\ 0 & 0 & x \\ 0 & 0 & x \end{bmatrix},$$

where "x" denotes a parameter to be estimated and "0" denotes a parameter fixed to zero. The scales of the factors are specified by standardizing ξ to unity, that is, the covariance matrix of ξ, Φ, is specified to have the pattern

$$\begin{bmatrix} 1 & & \\ x & 1 & \\ x & x & 1 \end{bmatrix},$$

TABLE 6.

Correlation matrix for Holzinger and Swineford Data (N = 145).

x_1	1.000								
x_2	0.318	1.000							
x_3	0.436	0.419	1.000						
x_4	0.335	0.234	0.323	1.000					
x_5	0.304	0.157	0.283	0.722	1.000				
x_6	0.326	0.195	0.350	0.417	0.685	1.000			
x_7	0.116	0.057	0.056	0.203	0.246	0.170	1.000		
x_8	0.314	0.124	0.229	0.095	0.181	0.113	0.585	1.000	
x_9	0.489	0.239	0.361	0.309	0.345	0.280	0.408	0.512	1.000

x_1 = Visual Perception x_2 = Cubes
x_3 = Lozenges x_4 = Paragraph Comprehension
x_5 = Sentence Completion x_6 = Word Meaning
x_7 = Addition x_8 = Counting Dots
x_9 = Straight-Curved Capitals

TABLE 7.

Modification indices for Λ.

	Visualization	Verbal intelligence	Speed
x_1	0	0.266	3.949
x_2	0	0.664	0.974
x_3	0	0.032	1.357
x_4	0.004	0	0.683
x_5	0.342	0	2.050
x_6	0.275	0	0.308
x_7	10.858	0.148	0
x_8	2.607	9.870	0
x_9	24.643	9.881	0

Modification indices for Θ

x_1	0.000								
x_2	0.631	0.000							
x_3	1.833	4.365	0.000						
x_4	0.041	0.741	0.044	0.000					
x_5	0.008	1.293	0.630	0.168	0.000				
x_6	0.013	0.129	1.399	0.121	0.003	0.000			
x_7	4.190	0.423	4.561	0.601	0.860	0.084	0.000		
x_8	0.379	0.164	0.000	3.746	0.212	0.193	24.966	0.000	
x_9	9.081	0.020	1.031	0.341	0.414	0.019	3.905	8.577	0.000

where "1" denotes a parameter which is fixed equal to 1. The covariance matrix of $\boldsymbol{\delta}$, $\boldsymbol{\Psi}$, is specified to be a diagonal matrix with elements in the diagonal to be estimated. A maximum likelihood confirmatory factor analysis results in a goodness of fit measure indicating that the model is not tenable, χ^2 with 24 degrees of freedom equals 52.6. Inspecting the MI's as given in Table 7 reveals that the element (8,7) of $\boldsymbol{\Psi}$ may not be equal to zero, the MI is equal to 24.966, indicating that the χ^2 for the model will decrease by this amount if that parameter is set to be free. The addition of this parameter results in a model with 23 degrees of freedom and a χ^2 equal to 28.8, which corresponds to a probability level of 0.19. An interpretation of this added covariance would be that the tests Addition and Counting Dots are measuring a factor not covered by the other tests, for example, a factor of numerical ability. Another modification would be to let element (9,1) of Λ be free, since this MI is also large, and the difference between the MI's is negligible. The χ^2 for the resulting model is 29.0, so the choice among the models is rather a matter of interpretability. The latter model says that the test Straight-Forward Curves is not only a measure of Speed but also of Visualization.

If the MI's are compared with the first order derivatives, as given in Table 8, we see that the patterns within matrices are about the same, but that the largest first order derivative is found for element (9,1) of $\boldsymbol{\Psi}$. Freeing this element will reduce χ^2 to 42,47, which is considerably higher than what was achieved by using the MI's.

In order to illustrate the scale-dependence of first order derivatives we could, for example, multiply the ninth observed variable by a constant and compute the first order derivatives again. These would then be divided by the same constant in the ninth row of Λ. Thus, the first order derivatives are scale-dependent. For the MI's, however, there is no change when the observed variables are rescaled. The first order derivatives in a row of Λ will be divided by the constant and the expected second order derivatives will be divided by the square of the same constant, which in turn will imply that the MI's are unaffected, as seen from (8).

TABLE 8.

First order derivatives for Λ.

	Vizualization	Verbal intelligence	Speed
x_1	0	-0.028	-0.108
x_2	0	0.046	0.058
x_3	0	-0.010	0.063
x_4	-0.005	0	0.085
x_5	0.046	0	-0.145
x_6	-0.041	0	0.056
x_7	0.202	-0.032	0
x_8	0.091	0.250	0
x_9	-0.303	-0.261	0

First order derivatives for Θ

x_1	0								
x_2	0.067	0							
x_3	0.082	-0.174	0						
x_4	-0.031	-0.127	0.033	0					
x_5	0.013	0.161	0.120	-0.029	0				
x_6	-0.017	0.051	-0.178	0.025	0.004	0			
x_7	0.241	0.072	0.253	-0.127	-0.145	0.045	0		
x_8	-0.075	0.048	0.002	0.337	-0.077	0.073	-0.282	0	
x_9	-0.359	-0.016	-0.121	-0.096	-0.102	0.022	0.141	0.157	0

Bollen (1980) discusses a factor analysis model for the measures in Table 9. He starts with a one-factor solution with uncorrelated errors and finds that this model is not tenable. On theoretical grounds the model was modified to include some correlations among the errors, and on the basis of the estimates of this modified model he added some equality constraints across factor loadings. He found that the factor scores obtained from the final model correlated 0.979 with a simple-sum index of the variables. In order to illustrate the behavior of the MI's when equality constraints are inherent, the simple model

$$x_i = \lambda \xi + \delta_i,\ i = 1, 2, \ldots, 6, \tag{16}$$

where λ is a scalar was used as a starting model. The δ's were specified to be uncorrelated and to have equal variances. The variance of ξ was set fixed equal to 1. The

TABLE 9.

Covariance matrix of Six Components of Political Democracy (N=113).

x_1	100.5					
x_2	104.2	182.2				
x_3	45.4	66.7	33.1			
x_4	80.2	90.4	40.6	106.5		
x_5	86.7	103.3	48.2	84.4	197.7	
x_6	86.7	121.3	49.3	81.1	108.2	127.7

x_1 = Press Freedom
x_2 = Freedom of Group Opposition
x_3 = Government Sanctions
x_4 = Fairness of Elections
x_5 = Executive Selection
x_6 = Legislature Selection

TABLE 10.

Steps taken to modify the model for Political Democracy

Step	Parameter to free	Mod. index	Chi-square	d.f.	Prov. level
0	λ_3	103.14	177.39	14	0.000
1	Ψ_{32}	24.55	60.17	13	0.000
2	Ψ_{62}	11.25	33.37	12	0.000
3	Ψ_{65}	8.92	22.35	11	0.022
4	λ_2	5.82	13.45	10	0.199
5	Ψ_{63}	2.85	7.28	9	0.608

model in (16) is a factor model where all factor loadings are constrained to be equal. Using the MI's to modify the model results in the steps listed in Table 10. The final model is equivalent to that obtained by Bollen, except that λ_4 is constrained to be equal to λ_1, λ_5, and, λ_6. If this constraint is relaxed a model with 8 degrees of freedom and a χ^2 equal to 5.78 is obtained. However, the decrease in χ^2 is not significant.

Interdependent Systems Example

Right from the beginning it must be emphasized that this example is concerned with a *misuse* of modification indices. The example is chosen as an illustration of how the MI's perform in a situation when very little information is included in a model. In general MI's should be used in situations when all information available has been included in the model and when this information is not enough to obtain an acceptable fit of the model to data. The MI's would then be used as guidelines for modifying the model.

A general form of an interdependent linear system is

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \boldsymbol{\zeta}, \tag{17}$$

where $\mathbf{y}$ is a vector of observed variables, $\mathbf{B}$ a matrix of structural parameters and $\boldsymbol{\zeta}$ a vector of residuals.

Consider the data in Table 11. These data consist of a subset of data described in Keeves (1972) and used by Lohnes (1979) in the discussion of his factorial modeling procedure. To start the MI procedure we specify $\mathbf{B}$ in (17) to be a zero matrix and that the covariance matrix of $\boldsymbol{\zeta}$ is diagonal. From a substantive view we can impose some constraints on the $\mathbf{B}$ matrix, since any paths from y_3, y_4, y_5, or y_6 to y_1 or y_2 would be

TABLE 11.

Correlation matrix (N = 235).

y_1	1.000					
y_2	0.128	1.000				
y_3	0.178	0.177	1.000			
y_4	0.044	-0.080	0.060	1.000		
y_5	0.758	0.202	0.228	0.091	1.000	
y_6	0.310	0.208	0.235	0.315	0.385	1.000

y_1 = 1968 Science Achievment y_2 = 1968 Science Attitude
y_3 = 1969 Science Class Processes y_4 = 1969 Peer SciMath Activities
y_5 = 1969 Science Achievment y_6 = 1969 Science Attitude

TABLE 12.

Steps taken to modify model for an Interdependent System

Step	Parameter to free	Mod. index	Chi-square	d. f.	Prov. level
0	β_{51}	134.45	306.04	15	0.000
1	β_{65}	34.68	106.05	14	0.000
2	β_{64}	21.53	68.51	13	0.000
3	β_{62}	7.56	45.72	12	0.000
4	β_{31}	7.41	37.62	11	0.000
5	β_{52}	6.06	30.09	10	0.001
6	β_{32}	5.75	23.84	9	0.005

uninterpretable. Thus, any suggestion by the procedure to include these paths should be ignored. It should be clear, that the procedure can not distinguish between adding the path y_5 to y_1, say, from adding the path y_1 to y_5, since all information available are the covariances among the variables and those covariances do not contain any information concerning possible directions of influences. In Table 12, the steps taken by the procedure are listed. The resulting model is depicted in Figure 2. This model could be compared with a model hypothesized by theory in Lohnes and differs from the model in Figure 2 in that the paths y_2 to y_5 and y_5 to y_6 are deleted and the paths y_1 to y_6 and y_3 to y_6 are added. This model has a χ^2 equal to 27.02 with 9 degrees of freedom, as compared with 23.84 with the same number of degrees of freedom for the model in Figure 2. Thus, without any a priori specifications, the procedure is able to find a model that is close to the theoretical model.

Summary and Conclusions

Anyone performing a statistical analysis should be aware that the statistical method can not do more than it is intended to do, and it is important for the analyst to be aware of the limitations of the method. In the area of model modification this is more important than ever. One should be aware that by adding parameters to a model, one can always come up with a model that fits any data perfectly. Of course, the result of such an analysis is of little value. The procedures described in this paper are intended for situations where the model is specified on the basis of substantive theory and includes as much as possible of what is known about the data to be analysed. If this

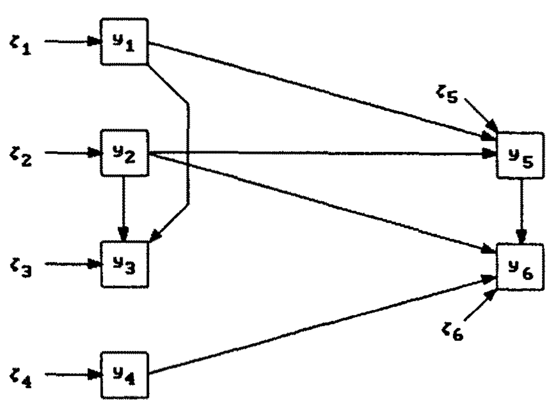

FIGURE 2
Model resulting from "automatic modification".

initial model does not fit the data well, the modification indices described in this paper may be useful instruments for detecting misspecifications in the model and guidelines in determining what to do to make the model fit the data better. It should be emphasized that at each step of the procedure, the analyst must use all his skill and all his information of the data and only accept models that are relevant, meaningful and interpretable.

Suppose extensive resources have been spent on data collection and all possible efforts have been laid down on formulating a model, but the data analysis indicates that the model does not fit. Rather than just accept this fact and leave it at that, it makes more sense to modify the model so as to fit the data better. It should also be clear, however, that since the model is modified and the data reanalyzed, the ordinary statistical hypothesis testing theory is no longer applicable. If the model is modified on the basis of the data, one is in reality performing an exploratory analysis and the findings of such an analysis should be subjected to a confirmatory analysis of new data.

References

Aitchison, J., & Silvey, D. C. (1958). Maximum likelihood estimation of parameters subject to restraints, *Annals of Mathematical Statistics*, *29*, 813–828.

BMDP (1977). *Biomedical computer programs*. Los Angeles: University of California Press.

Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review*, *45*, 370–390.

Costner, H. L., & Schoenberg, R. (1972). Diagnosing indicator ills in multiple indicator models. In A. S. Goldberger & O. D. Duncan (Eds.), Structural education models in the social sciences (pp. 167–199). New York: Seminar Press.

Dijkstra, T. K. (1981). *Latent variables in linear stochastic models*. Unpublished doctoral dissertation, University of Groningen.

Herting, J. R., & Costner, H. L. (1985). Respecification in multiple indicator models. In H. M. Blalock, Jr. (Ed.), *Causal models in the social sciences* (2nd ed., pp. 321–393). Hawthorne, NY: Aldine.

Holzinger, K., & Swineford, F. (1939). A study in factor analysis: The stability of a bifactor solution. *Supplementary Education Monograph No. 48*. Chicago: University of Chicago Press.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183–202.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V Analysis of linear structural relationships by maximum likelihood and least squares methods*. Chicago: International Educational Services.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI Analysis of linear structural relationships by the method of maximum likelihood*. Mooresville, IN: Scientific Software.

Keeves, J. P. (1972). *Educational environment and student achievement* (Stockholm Studies in Educational Psychology No. 20). Stockholm: Almqvist & Wiksell.

Lohnes, P. R. (1979). Factorial modeling in support of causal inference. *American Educational Research Journal*, *16*, 323–340.

MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107–120.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, *44*, 50–57.

Saris, W. E., de Pijper, W. M., & Zegwaart, P. (1979). Detection of specification errors in linear structural equation models. *Sociological Methodology 1979*.

Sörbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology*, *28*, 138–151.

Manuscript received 10/29/86
Final version received 3/21/88