4

THE DETECTION AND CORRECTION OF SPECIFICATION ERRORS IN STRUCTURAL EQUATION MODELS

Willem E. Saris

UNIVERSITY OF AMSTERDAM

Albert Satorra

UNIVERSITY OF BARCELONA

Dag Sörbom

UNIVERSITY OF UPPSALA

The authors thank Brendan Bunting and several anonymous reviewers for their useful comments on an earlier version of this paper. This research was supported by a grant from the Spanish Ministry of Education and Science.

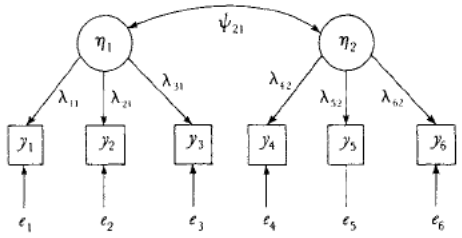
Detection of specification errors in structural equation models normally requires two steps. The first step establishes whether (further) corrections of the model are necessary, and the second detects the restrictions that have to be corrected. The χ^2 test is frequently used in the first step, even though this test has been severely criticized. The modification index introduced by Jöreskog and Sörbom in the LISREL program is commonly used in the second step. In this paper, we show that both procedures have serious problems and that the suggestions for corrections are not necessarily correct. Both procedures are affected not only by the size of the misspecification in the model but also by the kind and position of the misspecification and by other characteristics of the model. Thus, some small errors in the model may be detected and corrected while large errors go undetected. We propose a new procedure to overcome this problem.

The detection and correction of specification errors in structural equation models normally requires two steps. In the first step, one determines whether corrections are necessary. The model test in programs like LISREL (Jöreskog and Sörbom 1983) or EQS (Bentler 1985) is commonly used in this step. In the second step, corrections are made to the model. Several procedures have been used to detect and correct specification errors in structural equation models (see Costner and Schoenberg 1973; Sörbom 1975; Saris, de Pijper, and Zegwaart 1979; Jöreskog and Sörbom 1983), but one of the most common procedures uses the modification index (MI) computed by LISREL (Jöreskog and Sörbom 1983).

In this paper, we show that these procedures, although very popular, do not necessarily lead to the best correction from a substantive point of view. The test statistic and the MI are more sensitive to misspecification in some restricted parameters than in others. Therefore, they may suggest corrections for parameters that deviate only slightly from the hypothesized values but not for parameters that are far from the hypothesized values.

We start with a brief introduction to the likelihood-ratio test commonly used to evaluate structural equation models. Next, we introduce the procedure for computing the power of the test and show that the power varies with the kind of error and the characteristics of the model. From this result, it follows that the MI also varies with the characteristics of the model. Then, we analyze the problem and suggest an alternative procedure for determining the important misspecifications in the model. Finally, we discuss the results obtained.

FIGURE 1. A measurement model.



THE LIKELIHOOD-RATIO TEST

In our approach, a structural equation model implies that the covariance matrix Σ of the observed variables is expressed as a function of a vector containing the fixed and free model parameters. That is,

$$\Sigma = \Sigma(\pi), \tag{1}$$

where π is a partitioned vector with q parameters (π_f), for which no value is specified a priori, and a vector with p parameters (π_r), for which the values are restricted a priori.[1]

Although this approach can be used for any type of structural model, with or without latent variables, we use as an example a very simple measurement or factor-analysis model. A path diagram of the model is presented in Figure 1. In this model, there are only two latent variables (η_1 and η_2) and six observed variables (y_1 to y_6). It is assumed that each latent variable affects only three observed variables and that the error terms are uncorrelated with each other and with the latent variables.

The model shown in Figure 1 is a specific case of the more general class of factor-analysis models for which the following model specification holds:

$$\Sigma = \Delta\Psi\Delta' + \Theta, \tag{2}$$

where Σ is the variance-covariance matrix for the observed variables, Ψ is the variance-covariance matrix for the latent variables, Θ is the

[1] By a convenient (local) reparameterization, this formulation includes the model specified in (1) with more general types of constraints on the parameters, like $h(\pi) = 0$, where $h(\cdot)$ is a "well-behaved" vector valued function of π.

variance-covariance matrix for the measurement errors, and Δ is the matrix of the loadings.

The model in Figure 1 specifies the following restrictions in the matrices mentioned in equation (2):

$$\Delta = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}, \quad \Psi = \begin{bmatrix} 1 & \psi_{21} \\ \psi_{21} & 1 \end{bmatrix},$$

$$\Theta = \begin{bmatrix} \theta_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{66} \end{bmatrix}. \tag{3}$$

In these three matrices, some parameters have been given fixed values. These specifications represent restrictions on the general model specified in equation (2). The values of these parameters are the elements of the vector π_r of equation (1). Other parameters have been represented by a symbol. These parameters are the q free parameters (π_f) that have to be estimated.

Misspecifications can exist for all the restrictions introduced in the model. Thus, in this model, we can distinguish three different kinds of misspecifications:

1. An omitted effect of a latent variable on an observed variable (for example, it is possible that $\lambda_{41} \neq 0$)
2. An omitted correlation between the error terms of observed variables of one latent variable (for example, $\theta_{21} \neq 0$)
3. An omitted correlation between the error terms of observed variables of two different latent variables (for example, $\theta_{41} \neq 0$)

It is common practice to estimate the parameters of a model, as specified in (2) and (3), by minimizing the following function (4) with

respect to the free parameters in π:

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\pi)) = \ln|\boldsymbol{\Sigma}| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \ln|\mathbf{S}| - m, \tag{4}$$

where $\mathbf{S}$ is the usual sample covariance matrix and m is the number of observed variables.

If the observed variables have a multivariate normal distribution, it can be shown that

$$-(n/2)F(\mathbf{S}, \boldsymbol{\Sigma}(\pi)) = \ln L(\pi), \tag{5}$$

where $n = (N-1)$, N is the sample size, and $\ln L(\pi)$ is the corresponding log-likelihood function except for an additive constant. Thus, the minimization of (4) provides maximum likelihood estimates for the free parameters in π. It can also be shown that the likelihood-ratio test statistic corresponding to the test of the specified null model against a less restrictive alternative model is

$$T = n(F_0 - F_a), \tag{6}$$

where F_0 is the minimum of the function (4) for the specified null model and F_a is the minimum of the function for the more general alternative model (Jöreskog 1981).

In general, we can use this test statistic to test the extra restrictions that have been introduced in the null model. If, however, the alternative model is a just-identified model that does not impose any restrictions on the covariance or correlation matrix, then $F_a = 0$ and $T = nF_0$. This test statistic is used as an overall goodness-of-fit statistic or as a model-test statistic for the null model. A more elaborate discussion of this point can be found in Matsueda and Bielby (1986).

If the null model under scrutiny is correct and the observed variables have a multivariate normal distribution, then the test statistic T will be asymptotically distributed as a χ^2 variable with degrees of freedom (df) equal to the number of independent elements in the variance-covariance matrix minus the number of model parameters that have to be estimated. On the basis of this result, we can test structural equation models as follows: The model of interest is rejected if

$$T > C_\alpha, \tag{7}$$

where C_α is the value of the χ^2 variable for which

$$\mathrm{pr}\left(\chi^2_{df} > C_\alpha\right) = \alpha. \tag{8}$$

In this formulation, C_α is the critical value of the test and α is the (nominal) significance level. In general, α is chosen to be 0.05. Given this choice, the critical value can be obtained from the χ^2 table.

THE OVERALL MODEL TEST RECONSIDERED

Although commonly used, the test described above is not without serious limitations. A frequent criticism is that the sample size influences the test statistic (e.g., Wheaton et al. 1977; Jöreskog and Sörbom 1983; Bentler and Bonett 1980; Hoelter 1983). However, a more general problem, indicated by Saris and Stronkhorst (1984), Satorra and Saris (1985), and Matsueda and Bielby (1986), is that this testing procedure completely ignores the power of the test, which is defined as the probability that an incorrect model will be rejected.

If in a testing situation the test statistic is significant, one tends to reject the model. However, if there is a high probability that a small specification error will lead to a significant value of T (high power), the decision to reject the model is no longer justified. The significant value of the test statistic can be due to a large or a small specification error in the model.

The opposite situation can also arise. If the test statistic is not significant, one tends not to reject the model. However, if the power is low, the testing result does not preclude the existence of large specification errors in the model.

These two examples clearly show that information about the power of the test is necessary before one can evaluate the results of the test. For a more elaborate discussion of this point, we refer to Saris and Stronkhorst (1984) and Matsueda and Bielby (1986). Given this requirement, we will discuss the evaluation of the power of the test.

It can be shown that for an incorrect but not highly misspecified model, the test statistic T is asymptotically distributed as a noncentral χ^2 variable with noncentrality parameter λ, which can be computed according to the procedure outlined below (Satorra 1983; Satorra and Saris 1983, 1985).

The noncentral χ^2 distribution with df degrees of freedom and noncentrality parameter λ will be denoted by $\chi^2_{df}(\lambda)$. Once λ has been determined, the power of the test can be obtained using various programs or from the tables for the $\chi^2_{df}(\lambda)$ in the following way

(Haynam, Govindarajulu, and Leone 1973):

$$\text{power} = \text{pr}\left(\chi^2_{df}(\lambda) > C_\alpha\right). \tag{9}$$

Satorra and Saris (1985) show that if the values of the parameters of the specified model are approximately known, the power of the model test for specified deviations from the model can be calculated as follows:

1. Choose the type of misspecification to be detected and the size of the deviations from the tested hypothesis.
2. Compute the covariance or correlation matrix with the values of the parameters of the original model and the values of the additional parameters, which are specified in step 1. The values of these additional parameters should be chosen to be equal to the hypothesized value plus the deviation to be detected.
3. Analyze this population variance-covariance matrix with the model to be tested. This analysis will provide a value of the test statistic. The deviation of the test statistic from zero can only be due to the misspecification introduced in the model in steps 1 and 2. It can be shown that the value of the test statistic in this population study is equal to the noncentrality parameter λ (Satorra and Saris 1985).
4. Using the value of λ obtained in step 3, the *df* corresponding to the model tested, and the significance level (α) of the test, determine the power of the test from the tables for the noncentral χ^2 distribution or with any standard program using equation (9).

Alternative procedures for computing the noncentrality parameter are discussed by Satorra and Saris (1985) and Matsueda and Bielby (1986), but Satorra and Saris argue that the procedure suggested above is the simplest to use and can always be applied. Satorra, Saris, and de Pijper (1985) compared the different procedures and found that asymptotically, the procedures are identical but that the suggested procedure gives the most accurate approximation of the power for small samples.

The procedure outlined above can be used to obtain the power of the χ^2 test for different kinds of specification errors. Matsueda and Bielby (1986) use this procedure to show that the size of the loadings

TABLE 1

The Values of the Noncentrality Parameters and the Power of the χ^2 Test for the 2×3 Factor-Analysis Model in Figure 1 ($N = 500$)

Type of Misspecification	Estimated Using Satorra-Saris Procedure		Estimated Using Modification Index	
	λ	Power	λ	Power
$\lambda^s_{41} = 0.1$ but is assumed to be 0.0	17.90	0.87	18.13	0.87
$\theta^s_{21} = 0.1$ but is assumed to be 0.0	0.07	0.05	0.07	0.05
$\theta^s_{41} = 0.1$ but is assumed to be 0.0	2.56	0.16	2.49	0.16

Note: s = standardized.

and the number of indicators have an effect on the power of the test. In a more elaborate study, Saris and Satorra (in press) discuss the effects of several characteristics of the models. In this paper, we show only one of these effects, namely, that the likelihood-ratio test has unequal power for misspecifications of the same size in different places in the model.

We applied this procedure to the model in Figure 1, introducing the three mentioned types of errors separately. In these computations, the size of the omitted correlation is always 0.1, and the variances of the latent variables are always 1, which means that the factors are standardized. The correlation between the factors is always 0.3. We set the loadings (λ_{ij}) equal to 0.9 and the measurement error variance (θ_{ii}) equal to 0.19. Given that we wanted to introduce the same misspecification error in all cases, we set λ_{41} equal to 0.1 and the values of θ_{21} and θ_{41} equal to 0.19, because these coefficients represent covariances instead of correlations.

The results of the power computations using the above discussed procedure are presented in Table 1. This table shows that the test is not equally sensitive to the different kinds of specification errors, even when these errors are the same size. The test has a high power for the λ_{41} coefficient but does not appear to be very sensitive to the other two errors.

Although we have demonstrated this phenomenon for only a specific type of model, the result has a wider significance. The above

result can occur for all types of models, but it is very difficult to predict the consequences for specific models.

We can conclude that an overall goodness-of-fit test for a model is questionable. This example illustrates that the same test is unequally sensitive to different kinds of specification errors, which means that only some errors will be detected. Moreover, there is no way to adjust the test to improve this situation. If we try to improve the power of the test for the correlated errors, the power for the loadings, like λ_{41}, will become so large that any small deviation will lead to rejection of the model.

Thus, we conclude that the model test, being a test for a multiple hypothesis, can give us only limited information about the quality of the model. If the model is not rejected, it does not mean that the specification errors in the model are small. We can only conclude that the specification errors for the parameters for which the test is sensitive are small. For the other parameters, large deviations are still possible. Similarly, if the model is rejected, we cannot say that it necessarily contains large misspecification errors. It is also possible that small deviations exist for parameters for which the test is very sensitive.

Only a very elaborate evaluation of the power of the test for the different kinds of specification errors in the model could give information from which conclusions could be drawn. Given all these complications, one might conclude that other procedures should be recommended.

DETECTING MISSPECIFIED RESTRICTIONS

In the previous section, we saw that the multiple hypothesis characteristic of the model test makes it difficult to control for statistical power. In this section, we discuss procedures to detect which restriction in the null model is misspecified. All the single restrictions are evaluated separately, even though this might be a simplification of reality.

To introduce these procedures, let us look once more at the problem we are trying to solve. The analyzed null model (H_0) may be seen as a restricted model of a more general (or alternative) model, say H_a. The alternative model (H_a) can be any model consisting of H_0 "augmented" by dropping one of its implied restrictions (e.g., by

adding one fixed parameter to the free parameters). Obviously, there are many models H_a that can be chosen as more general or alternative models of H_0. If the validity of H_0 is questioned a priori, and if there is no prior information about which of the restrictions implied by H_0 is incorrect, then the problem of choosing a model H_a as an alternative to H_0 arises.

One way to choose between the different restrictions to be relaxed is to use the statistic T, given in (6). This statistic can be computed for each of the possible alternative models (H_a), and the significance of each of these test statistics can be evaluated. The χ^2 distribution with one *df* can be used as the asymptotic null distribution of the test statistic. If the most significant (or the largest) value of the test statistic is used to determine which restriction has to be relaxed, the restriction chosen would be the one that leads to the largest decrease in the χ^2 value. This method, however, requires heavy computations, because there are many H_a models to analyze; in fact, there are as many models as there are different single restrictions in the model. Therefore, one should look for alternative procedures in the literature.

The use of the first-order derivative of the log-likelihood function for the different fixed parameters has been suggested in the older literature (Byron 1972; Sörbom 1975; Saris et al. 1979). This procedure is based on the Lagrange multiplier test, which is based on the following argument.

Consider a parameter π_i, which the analyzed model H_0 restricts at the value π_0. Once H_0 is fitted, one can consider the value $d_{\pi_i} = \partial \ln L(\pi)/\partial \pi_i$ evaluated at the estimated values of the parameters obtained under the specified model H_0. Under the specification H_0, $\ln L(\pi)$ is maximized with respect to the free elements of π, and π_i is kept fixed at π_0. Thus, unless π_0 is the value of π_i at which the log-likelihood function gets its maximum, d_{π_i} will, in general, differ from zero, indicating that a shift of π_i from π_0 can increase the value of $\ln L(\pi)$. Thus, a large deviation of d_{π_i} from zero may invalidate the hypothesis that $\pi_i = \pi_0$.

A formal test can be specified for this statistic. Silvey (1959) has shown that in the case of a correct model and under some regularity conditions, the estimates of the first-order partial derivatives evaluated at the estimated values (which can be shown to be equal to the Lagrange multipliers) are multivariate normal distributed with expectation equal to zero if the restrictions are true. The variance-covariance

matrix of these estimates can be calculated from an estimate of the information matrix $\mathbf{Q}$, which is defined as

$$\mathbf{Q} = E\left[-\partial^2 \ln L(\pi)/\partial\pi\partial\pi'\right]. \tag{10}$$

When we partition this matrix $\mathbf{Q}$ into

$$\begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix},$$

where $\mathbf{Q}_{11}$ contains elements derived for the fixed parameters and $\mathbf{Q}_{22}$ contains elements derived for the free parameters, the variance-covariance matrix $\mathbf{V}$ for the estimates of the Lagrange multipliers can be proved (Silvey 1970) to be

$$\mathbf{V} = \mathbf{Q}_{11} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}. \tag{11}$$

For each of the restricted parameters, a standard normal test statistic can be evaluated using the quantity z_{π_i}:

$$z_{\pi_i} = \left(\partial \ln L(\pi)/\partial\pi_i\right)/v_i, \tag{12}$$

where v_i is the square root of the ith diagonal element of $\mathbf{V}$. Byron (1972) has suggested dropping all restrictions connected with significant z_{π_i}. Sörbom (1975) has suggested a stepwise relaxation of the restriction that is connected with the largest first-order derivative (d_{π_i}). Saris et al. (1979) have suggested a weighted procedure to take into account the interdependencies between the statistics.

Nowadays, the more commonly used procedure in structural modelling is the MI, routinely computed by LISREL (Jöreskog and Sörbom 1983). Although several forms of the MI have been given (see the appendix), the last version provides the score statistic. For individual parameters, this statistic (S_{π_i}) can be derived as follows (see the appendix):

$$S_{\pi_i} = z_{\pi_i}^2. \tag{13}$$

This means that a very simple relationship exists between the two statistics and that they will have similar characteristics with respect to the power of the test.

The asymptotic equivalence between the likelihood-ratio test and the corresponding score test is a classical result (see, e.g., Buse 1982). This asymptotic equivalence is shown in the appendix. It implies that the asymptotic distribution of the MI is the same as the

distribution of the difference statistic T, given in (6), when the model H_a is equal to the model H_0 augmented with the parameter for which the restriction is tested. Under standard conditions, this statistic is an asymptotic χ^2 statistic with 1 *df*.

Thus, the asymptotic distribution of the MI when H_0 is correct is χ^2 with 1 *df*, and the asymptotic distribution when H_0 is misspecified is noncentral χ^2 with 1 *df* and a specific noncentrality parameter, say λ. This noncentrality parameter will be the same as the noncentrality parameter for the statistic T. Given a specific alternative covariance matrix, λ can be approximated by taking the difference of the χ^2 values of H_0 and H_a obtained by analyzing the alternative specified population covariance matrix. Now we can say that the MI provides an alternative way of approximating λ; the power can be determined using as λ the corresponding MI obtained when the alternative population covariance matrix is analyzed. It can be shown that both approximations are asymptotically equal (Satorra 1986).

Table 1 shows the values of the noncentrality parameter calculated using the MI for the model presented in Figure 1. This table shows that the values of λ estimated with the MI for the different parameters are approximately identical to the noncentrality parameters computed using the power calculations given before. Thus, all the remarks made about the model test can also be made about the MI. Consequently, one can get very large values of the MI if this statistic is very sensitive for this parameter (high power), and one can get very low values of the MI for parameters for which this statistic is not very sensitive. Thus, the phenomena that can occur when the statistic T is used can also occur when the MI is used; i.e., large specification errors may remain undetected while small specification errors may be detected and corrected.

FIGURE 2. A structural equation model with standardized variables and parameters.

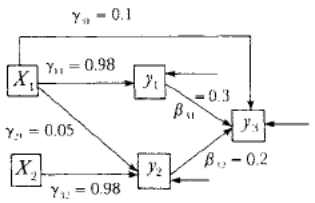


TABLE 2
The Population Correlation Matrix Computed from the Parameter Values Specified in Model 2, Assuming that the Variables Have a Variance of 1

	y_1	y_2	y_3	x_1	x_2
y_1	1.000				
y_2	0.0490	1.000			
y_3	0.4078	0.2197	1.000		
x_1	0.9800	0.0500	0.4040	1.000	
x_2	0.0000	0.9800	0.1960	0.0000	1.0000

Given the close relationship between the statistic T, the MI, and the Lagrange multiplier, we must conclude that the use of all three procedures can lead to corrections in the model that are not necessary and to omissions of corrections that are necessary. We will illustrate this point once more using the MI.

So far, we have discussed only the power and the MI for parameters in different models, but the same characteristics can be found for several parameters within the same model. To illustrate this point, we give a different example. The model is presented in Figure 2.

The values of the parameters indicated in Figure 2 have been used to compute the population covariance matrix of the standardized variables for this model. This correlation matrix is presented in Table 2. This population covariance matrix is analyzed using the model

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \Gamma\mathbf{x} + \zeta, \tag{14}$$

assuming that $E(\mathbf{y}) = \mathbf{0}$, $E(\mathbf{x}) = \mathbf{0}$, $E(\zeta) = \mathbf{0}$, $E(\xi\xi') = \Psi$ is diagonal, and $E(\mathbf{x}\zeta') = \mathbf{0}$, where $\mathbf{B}$ and Γ are specified as follows:

$$\mathbf{B}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \beta_{32} & 0 \end{bmatrix} \qquad \Gamma = \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \\ \gamma_{31} & 0 \end{bmatrix}. \tag{15}$$

Comparing this null model with the model that has been used to generate the data, we can see that except for the parameters β_{31} and γ_{21}, which have been fixed at zero, the same parameters are specified. This means that a small specification error is introduced in the null model for the parameter γ_{21} and a big specification error is introduced for the parameter β_{31}.

When this matrix is analyzed with the specified model, we find that the MIs for γ_{21} and β_{31} are 6.250 and 0.443. This means that the parameter for which the smallest misspecification is made is denoted as the parameter with the largest error, while a serious error of 0.30 in the model for β_{31} goes undetected. The reason for this result is the same as the reason given above. It can be shown that the power of the test for the restriction associated with the first parameter is much higher than the power of the test for the restriction associated with the second parameter. This example clearly shows that the MI cannot be used as simply as it usually is. At the least, one should expect that the use of this index can lead to incorrect conclusions in many instances. Therefore, a different approach is needed. We will discuss this approach in the next sections. Let us first look more precisely into the problem.

WHAT IS THE PROBLEM?

The problem with the use of the MI is that one is not taking into account the variation in the sensitivity of the index to different types of restrictions and to the characteristics of the model. In the example above, we created a situation in which the MI is much more sensitive for one specification error than for another. We could create this situation because we have found in earlier studies (Saris and Satorra, in press) a relationship between the power of the test and the explained variance in the equation. In an equation with a high explained variance, the power will be high. In an equation with a low explained variance, the power will also be low. This principle has been applied here.[2]

The analysis suggests that one should consider not only the MI but also the sensitivity of the MI to specification errors of different sizes for different restricted parameters. Such a sensitivity analysis requires the calculation of the noncentrality parameter for several parameters and for different sizes of the specification errors. We have seen that the noncentrality parameter is equal to the value of the MI for population data.

[2] This contradictory result could not be explained by the position of the errors in the model, as one reviewer suggested. If the order of the error changes, the same effect occurs.

TABLE 3
The Modification Indices for the Model in Figure 2

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0.42 \\ 6.00 & 0 & 1.05 \\ 0.44 & 0 & 0 \end{bmatrix}$$

$$\mathbf{\Gamma} = \begin{bmatrix} 0 & 0 \\ 6.25 & 0 \\ 0 & 0 \end{bmatrix}$$

The noncentrality parameter can be calculated using the procedure discussed above, although this procedure is rather tedious for large models. Also, one should be careful in interpreting the estimated values of λ if the specification errors are very large (Satorra et al. 1985).

Table 3 provides the results of these calculations for the example discussed above. In this table, the MIs are presented for a correlation matrix from a population for which the model of Figure 2 holds, and the data are analyzed with a model in which β_{31} and γ_{21} are zero. Table 4 presents the MIs for different sizes of the misspecifications.

Table 3 shows that two parameters, γ_{21} and β_{21}, have relatively large MIs and that the other parameters have small MIs. On the other hand, we see in Table 4 that the MI increases very rapidly for γ_{21} and β_{21} but that it is still very small for a specification error of 0.1 in parameter β_{31}. The latter phenomenon explains how the last parameter can have a small MI and an error considerably larger than the errors for the other parameters.

TABLE 4
The Modification Indices for Different Sizes of the Misspecifications

	Size of the Misspecification								
Parameter	0.01	0.03	0.05	0.07	0.09	0.1	0.2	0.3	0.4
γ_{21}	0.25	2.30	6.52	13.20	22.88	—	—	—	—
β_{21}	0.25	2.30	6.52	13.20	22.88	—	—	—	—
β_{31}	0	0	0	0	0	0.06	0.26	0.73	1.87

This information could be used to determine the likely size of the misspecification for each fixed parameter. This result can be obtained by comparing the values of the MI with the obtained value. In Table 4, we see that the specification error for γ_{21} and β_{21} is probably 0.05, which is correct for γ_{21}, and that the specification error for β_{31} is between 0.2 and 0.3, which is a little bit too low.[3] But the conclusion from this analysis—that the last parameter probably deviates the most from the hypothesized value—is correct, and it contrasts sharply with the conclusion that would have been drawn if only the MI had been used. This analysis suggests that one should concentrate on the size of the misspecification itself instead of on the size of the MI to determine where the model should be corrected

A NEW PROCEDURE FOR DETECTING SPECIFICATION ERRORS

In the previous section, we outlined a procedure to determine the parameter that probably deviates the most from its population value. However, this procedure is very tedious, because one has to calculate the power functions for many parameters.

Fortunately, it is also possible to obtain a direct estimate of the size of the misspecification for the restricted parameters. Assuming that H_0 restricts a parameter π_i at the value π_0, the change in the value of the parameter when this parameter is added as a free parameter to H_0 may be approximated as

$$\pi_i - \pi_0 = \mathrm{MI}/d_{\pi_i}, \tag{16}$$

which can be calculated, for any fixed parameter, from the results of fitting H_0. This approximation is justified in the appendix, where the MI is the score statistic S_{π_i}. Note that because of the specific function minimized in LISREL, d_{π_i} is $-(N-1)$ times the corresponding first derivatives reported in the output of the program.

In practice, one may observe some deviation between the estimated parameter change and the observed parameter change when in fact model H_a is fitted because the result is only asymptotically correct. However, very often the approximation will be adequate for all

[3] This may be due to the presence of two misspecifications in model H_0 and to our examination of only one misspecification at a time.

practical purposes. From the equalities shown in the appendix, we can easily see that the approximation would turn into an equality if the approximations by quadratic functions were in fact equalities; this would imply, however, that the matrix $\mathbf{Q}$ was constant for the range of values of the parameter between π_i and π_0.

One of the advantages of using this statistic is that the information for all the restricted parameters can be obtained by using model H_0 to analyze the data. This also means that all parameters that are restricted under H_0 may be evaluated in the same run in which the free parameters of the model are estimated. In programs like LISREL and EQS, these estimates are easily provided.

It is also possible to obtain an estimate of the standard error for this statistic, but we will not discuss that here. This information is already available if we have the value of the MI, since the MI is equal to the squared ratio of $\pi_i - \pi_0$ and the standard error. This means that taking into account the standard error would bring us back to the MI. Because the purpose of this new approach is to find an alternative to this statistic, we will not discuss the standard errors any further.

Table 5 presents the differences between the hypothesized values of the parameters and the values that can be expected if the parameters are left free for the above-mentioned example. These differences were calculated with equation (16). The table illustrates that in this example, the correct values are obtained for the two parameters under discussion. Thus, in this case, these estimates of the changes in the parameters provide the correct information, and one can draw the correct conclusion that β_{31} deviates the most from the hypothesized value and that γ_{21} deviates only minimally.

It should be mentioned, however, that the results of this procedure do not have to be that good all the time. First of all, one has to

TABLE 5
The Estimated Changes in the Parameters for the Model in Figure 2

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0.01 \\ 0.05 & 0 & 0.02 \\ 0.30 & 0 & 0 \end{bmatrix}$$

$$\mathbf{\Gamma} = \begin{bmatrix} 0 & 0 \\ 0.05 & 0 \\ 0 & 0 \end{bmatrix}$$

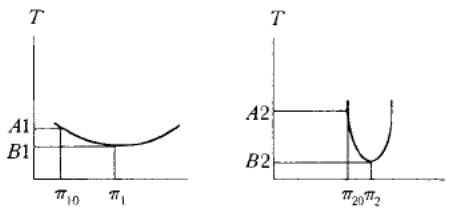
take into account random fluctuations in the data, which may lead to deviations in the estimates. Second, the derivation of equality (16) assumes that the misspecifications of H_0 are small. If this is not true, it might lead to biases in the estimates of the changes in the parameters. We will return to this point in the discussion.

DISCUSSION

In this paper, we have discussed three related procedures. In Figure 3, we have summarized the differences between these procedures. Figure 3 presents for two parameters the relationships between the fixed values of the parameters and the values of the fitting function used in a program like LISREL. The estimated values of the free parameters in the model are not fixed; they are allowed to change when the values of the fixed parameters studied in the graph also change. In general, the function studied is -2 times the log ratio of the maximum of the likelihood for the null model and any exactly identified model. This test statistic is asymptotically χ^2 distributed. Therefore, this figure also gives the changes in the value of this test statistic for changes in the value of the parameter. This is exactly the same kind of information we have provided in the sensitivity analysis.

In this figure, we have illustrated a situation in which the test statistic is much more sensitive to the deviation from the population value for one parameter (π_2) than for another parameter (π_1). This means that the model test is much more sensitive to deviations in the second parameter than to deviations in the first parameter. Because of this phenomenon, the model test does not do what it is meant to do. It

FIGURE 3. The relationship between two parameters and the values of T.



cannot provide a test for all the parameters of the model. It provides a test only for the parameters for which the test is sensitive. Therefore, we conclude that this test is only of limited use: If a model is not rejected, it does not mean that there are no large specification errors in the model. These errors are still possible for the parameters for which the test is not sensitive.

When the MI is used, we start with the null model. In the null model, some parameters have been given fixed values that might be wrong. The points $A1$ and $A2$ indicate the values of the test statistic for the misspecified values for two parameters. If these two parameters are made free parameters, the function will reach its minimum, assuming that this error was the only misspecification. On the vertical axes in Figure 3, we can see what the change in the test statistic will be for each of these changes in the model. For the first parameter, it will be $A1-B1$, and for the second parameter, it will be $A2-B2$. This information is given by the MI. We can also see that for the parameter with the more sensitive test, the MI and the change in the test statistic are much larger. On the other hand, it is also clear that this change is due to a much smaller misspecification in this parameter than in the other parameter. This result can be obtained by subtracting $\pi_{10} - \pi_1$ and $\pi_{20} - \pi_2$. In the last section, we suggest using the latter information, which is the expected change in the parameter, as an indication of where a correction might be introduced in the model.

In Table 6 we have indicated four possible situations, defined by the expected change in the parameter and the expected reduction in the test statistic, that can occur when this parameter is introduced in the model. If situation 1 occurs, it seems reasonable to introduce the

TABLE 6
Four Different Situations Defined by Expected Change in the Parameter and Expected Reduction in the Modification Index

Expected Reduction in the Modification Index	Expected Change in the Parameter π_i	
	Large	Small
Large	1	2
Small	3	4

parameter π_i in the model because it will lead to big changes in the values of the parameter and the test statistic. In the second situation, it does not make sense to introduce the parameter in the model, even though the reduction in the test statistic will be large, because the change in the parameter is expected to be very small. Also, if the change in the parameter and the reduction in the test statistic are expected to be small (situation 4), it does not make sense to introduce the parameter. But if the expected change in the parameter is large and the reduction in the test statistic is small (situation 3), then the situation is somewhat unclear. The large expected change might be due to sampling fluctuation, but it is also possible that the model test statistic is not sensitive for this parameter.

The discussion of these different situations shows that in general, the decision to introduce an extra parameter is based on the expected change in the parameter and not on the reduction in the test statistic. Whether this is a correct choice is open to debate, but it seems reasonable when we want to obtain values of the structural parameters that are as close as possible to the values in the population.

Although the new procedure seems to be a considerable improvement over those commonly used, some comment on this approach is needed. First, the correction of models on theoretical arguments is, of course, always better (Saris et al. 1979). No procedure can guarantee that substantively, the most important corrections are found. The result depends greatly on the model with which one starts. Furthermore, it should be recognized that the estimates obtained are sample statistics and therefore subject to sampling fluctuations. It also has to be said that all calculations are based on the assumption that the information matrix is approximately constant between π_0 and π_i. Thus, we cannot rely on these results when there is a very large misspecification present for any other parameter in the model.

The reader should also be aware that the estimated change in the restricted parameters depends upon the scale of the observed variables. This implies that when unstandardized variables are used, one has to be careful in interpreting the size of the parameter changes.

On a more theoretical plane, we note that the new procedure shifts our intention from obtaining a low test statistic to avoiding large deviations from the hypothesized parameter values. These are not necessarily the same. It is possible that when this approach is used, one ends up with a model that does not fit at the 0.05 level because the test

is very sensitive to substantively unimportant specification errors. However, use of the suggested approach gives us some assurance that large specification errors in the model have been detected.

Clearly, the above methodology can be extended to the case in which a generalized least squares discrepancy function is substituted for F in (4). In this way, asymptotic distribution-free methods (Browne 1982, 1984; Bentler 1983) can be included.

The above results can also be extended to the situation in which several restrictions are relaxed at the same time. Of course, the implementation and the practical implications of such an extension remain to be investigated.

APPENDIX

In this appendix, we show that the actual MI, which is a score statistic, is asymptotically equivalent to the reduction in the test statistic T if the model is augmented with one more parameter.

Suppose that π is the vector of free parameters of a model H_a and that some constraints are imposed on π by the more restricted model H_0. Suppose that $\hat{\pi}$ is the estimate of π when H_0 has been fitted. The score statistic corresponding to the constraints implied by H_0 on H_a takes the following form:

$$S = (\partial \ln L(\pi)/\partial\pi)'\mathbf{Q}^{-1}(\partial \ln L(\pi)/\partial\pi), \qquad \text{(A-1)}$$

where $\mathbf{Q} = E[-\partial^2 \ln L(\pi)/\partial\pi\partial\pi']$, and $\mathbf{Q}$ and $\partial \ln L(\pi)/\partial\pi$ are evaluated at $\hat{\pi}$ (e.g. Buse 1982).

Since the MI is just a score test of a single restriction, we consider the situation in which the parameter vector π' is partitioned as (π_i, π_f'), where by H_0, the parameter π_i is restricted to π_0, and π_f is the vector of free parameters. Then the following partitioning of $\mathbf{Q}$ is implied:

$$\mathbf{Q} = \begin{bmatrix} Q_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix}.$$

Because π_f is a vector of free parameters, $(\partial \ln L(\pi)/\partial\pi_f)$ evaluated at $\hat{\pi}_f$ equals the null vector; thus, $(\partial \ln L(\pi)/\partial\pi)$ partitions as $[d_{\pi_i}, \mathbf{0}']'$, where $d_{\pi_i} = \partial \ln L(\pi)/\partial\pi_i$ evaluated at $\hat{\pi}_f$. Using proper-

ties of the inverse of partitioned matrices, the score statistic of (A-1) corresponding to the restriction $\pi_i = \pi_0$, say S_{π_i}, takes the following form:

$$S_{\pi_i} = d_{\pi_i}^2 / \left(Q_{11} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\right). \tag{A-2}$$

This statistic is the MI used in LISREL VI. When restrictions are expressed as equality constraints, the above expression of the MI becomes more complex, but the same principle applies (see Sörbom 1986).

Another type of statistic arises by considering

$$S_{\pi_i}^* = d_{\pi_i}^2 / (Q_{11}). \tag{A-3}$$

This was, in fact, the MI implemented in earlier versions of LISREL.

Now, we derive the relationship between the above MIs and the decrease in the value of the model test statistic when models H_0 and H_a are analyzed. Given the covariance matrix $\mathbf{S}$ and the model H_0 augmented with the parameter π_i, the fitting function F given in (4) can be seen as a function of π_f and π_i; i.e., $F = F(\pi_i, \pi_f)$. By considering π_f to be fixed at $\hat{\pi}_f$, a Taylor series expansion of $F(\pi_i, \hat{\pi}_f)$ around the value π_i^* where the function $F(\pi_i, \hat{\pi}_f)$ attains its minimum and the equality $[\partial^2 F(\pi_i^*, \hat{\pi}_f)/\partial^2 \pi_i] = (2/n)Q_{11} + o(1)$, which holds when $|\pi_0 - \pi_i^*| = o(1)$, we obtain

$$F(\pi_i, \hat{\pi}_f) = F(\pi_i^*, \hat{\pi}_f) + (1/n)Q_{11}(\pi_i - \pi_i^*)^2 + o\left(|\pi_i - \pi_i^*|^2\right). \tag{A-4}$$

By differentiating both sides of the above equality with respect to π_i, we get

$$\partial F(\pi_i, \hat{\pi}_f)/\partial \pi_i = (2/n)Q_{11}(\pi_i - \pi_i^*) + o\left(|\pi_i - \pi_i^*|^2\right).$$

Then, giving to π_i the value of π_0, we get

$$(\pi_0 - \pi_i^*) = -d_{\pi_i}/Q_{11} + o\left(|\pi_0 - \pi_i^*|^2\right), \tag{A-5}$$

where the obvious equality $d_{\pi_i} = -(n/2)\partial F(\pi_0, \hat{\pi}_f)/\partial \pi_i$ has been used. Therefore, combining (A-4), $\pi_i = \pi_0$, and (A-5), we get

$$F(\pi_0, \hat{\pi}_f) - F(\pi_i^*, \hat{\pi}_f) = (1/n)d_{\pi_i}^2/(Q_{11}) + o\left(|\pi_i - \pi_i^*|^2\right).$$

That is,

$$nF(\pi_0, \hat{\pi}_f) - nF(\pi_i^*, \hat{\pi}_f) = S_{\pi_i}^* + o\left(n|\pi_0 - \pi_i^*|^2\right), \tag{A-6}$$

which suggests that the statistic $S_{\pi_i}^*$ is approximately equal to

$nF(\pi_0, \hat{\pi}_f) - nF(\pi_i^*, \hat{\pi}_f)$ or to the difference in χ^2 values when H_0 and H_a are estimated. In this approximation, we assume that $(n)^{1/2}(\pi_0 - \pi_i^*)$ is not large, which ensures that the remaining term of (A-6) is small. Note that in the above difference in χ^2 values, the parameters π_f remain fixed at $\hat{\pi}_f$, and π_i takes the value π_i^* for which the χ^2 function minimizes. It is clear, however, that because π_f is kept fixed at $\hat{\pi}_f$, in general, $S_{\pi_i}^*$ will underestimate the decrease in the χ^2 values when H_0 and H_a are in fact fitted.

Using the argument we used to derive (A-6), but considering now the function $G(\pi_i) = \min_{\pi_f} F(\pi_i, \pi_f)$ instead of $F(\pi_i, \hat{\pi}_f)$, we get

$$nF(\pi_0, \hat{\pi}_f) - nF(\tilde{\pi}_i, \tilde{\pi}_f) = S_{\pi_i} + o(n|\pi_0 - \tilde{\pi}_i|^2),$$

where, now, $\tilde{\pi}_f$ and $\tilde{\pi}_i$ are, respectively, the values of π_f and π_i obtained by fitting H_a, and S_{π_i} is the same as in (A-2). In this case, the χ^2 difference corresponds to the decrease that would be observed if both models were actually estimated. Note that $\tilde{\pi}_i$ is the value that minimizes $G(\pi_i)$. The expression $Q_{11} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}$, instead of Q_{11}, that appears in the expression of S_{π_i} arises as a result of having to consider the second derivative of the function $G(\pi_i) = \min_{\pi_f} F(\pi_i, \pi_f)$ and the following result of calculus:

$$\partial^2 G/\partial^2 \pi_i = \partial^2 F/\partial^2 \pi_i - \left(\partial^2 F/\partial \pi_i \, \partial \pi_f'\right)\left(\partial^2 F/\partial \pi_f \, \partial \pi_f'\right)^{-1} \times \left(\partial^2 F/\partial \pi_f \, \partial \pi_i\right),$$

which holds under very common assumptions imposed on the function $F(\pi_i, \pi_f)$. By considering $G(\pi_i)$, instead of $F(\pi_i, \hat{\pi}_f)$, we have approximated the decrease in the χ^2 value when π_i is made a free parameter and when π_f is allowed to shift from $\hat{\pi}_f$ as a result of incorporating a new free parameter in the model. Now the statistic S_{π_i} need not be smaller in value than the actual decrease in χ^2 values.

REFERENCES

BENTLER, P. M. 1983. "Some Contributions to Efficient Statistics in Structural Models: Specification and Estimation of Moment Structures." *Psychometrika* 48:493–517.

_____. 1985. *Theory and Implementation of EQS: A Structural Equations Program*. Los Angeles: BMDP.

BENTLER, P. M., AND D. G. BONETT. 1980. "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures." *Psychological Bulletin* 88:588–606.

BROWNE, M. W. 1982. "Covariance Structures." Pp. 72–141 in *Topics in Applied Multivariate Analysis*, edited by D. M. Hawkins. Cambridge: Cambridge University Press.

_____. 1984. "Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures." *British Journal of Mathematical and Statistical Psychology* 37:62–83.

BUSE, A. 1982. "The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note." *American Statistician* 36:153–57.

BYRON, R. P. 1972. "Testing for Misspecification in Econometric Systems Using Full Information." *International Econometric Review* 13:745–56.

COSTNER, H. L., AND R. SCHOENBERG. 1973. "Diagnosing Indicator Ills in Multiple Indicator Models." Pp. 167–99 in *Structural Equation Models in the Social Sciences*, edited by A. S. Goldberger and O. D. Duncan. New York: Seminar Press.

HAYNAM, G. E., Z. GOVINDARAJULU, AND F. C. LEONE. 1973. "Tables of the Cumulative Non-Central Chi-Square Distribution." Pp. 1–78 in *Selected Tables in Mathematical Statistics: Vol. 1*, edited by H. L. Harter and D. B. Owen. Providence: Mathematical Statistical Society.

HOELTER, J. W. 1983. "The Analysis of Covariance Structures: Goodness of Fit Indices." *Sociological Methods and Research* 11:325–44.

JÖRESKOG, J. W. 1981. "Analysis of Covariance Structures." *Scandinavian Journal of Statistics* 8:65–92.

JÖRESKOG, K. G., AND D. SÖRBOM. 1983. *LISREL V User's Guide. Analysis of Structural Relationships by Maximum Likelihood and Least Squares Methods*. Chicago: International Educational Services.

MATSUEDA, R. L., AND W. T. BIELBY. 1986. "Statistical Power in Covariance Structure Models." Pp. 120–58 in *Sociological Methodology 1986*, edited by Nancy B. Tuma. Washington, DC: American Sociological Association.

SARIS, W. E., W. M. DE PIJPER, AND P. ZEGWAART. 1979. "Detection of Specification Errors in Linear Structural Equation Models." Pp. 151–71 in *Sociological Methodology 1979*, edited by Karl F. Schuessler. San Francisco: Jossey-Bass.

SARIS, W. E., AND A. SATORRA. In press. "Characteristics of Structural Equation Models Which Affect the Power of the Likelihood Ratio Test." In *Sociometric Research*, vol. 2, edited by W. E. Saris and I. N. Gallhofer. London: Macmillan.

SARIS, W. E., AND L. H. STRONKHORST. 1984. *Causal Modelling in Nonexperimental Research: An Introduction to the LISREL Approach*. Amsterdam: Sociometric Research Foundation.

SATORRA, A. 1983. "Potència del contrast de la raó de versemblança en models d'equacions estructurals." Ph.D. diss., University of Barcelona.

_____. 1986. "Alternative Test Criteria in Covariance Structure Analysis: A Unified Approach." Research Report. Barcelona: University of Barcelona, Department of Statistics and Econometrics.

SATORRA, A., AND W. E. SARIS. 1983. "The Accuracy of a Procedure for Calculating the Power of the Likelihood Ratio Test as Used within the LISREL Frame-

work." Pp. 129–90 in *Sociometric Research 1982*, edited by C. O. Middendorp. Amsterdam: Sociometric Research Foundation.

_____. 1985. "The Power of the Likelihood Ratio Test in Covariance Structure Analysis." *Psychometrika* 50:83–90.

SATORRA, A., W. E. SARIS, AND W. M. DE PIJPER. 1985. *Several Approximations to the Power Function for the Likelihood Ratio Test in Covariance Structure Analysis*. Amsterdam: Sociometric Research Foundation.

SILVEY, S. D. 1959. "The Lagrange Multiplier Test." *Annals of Mathematical Statistics* 30:389–407.

_____. 1970. *Statistical Inference*. Middlesex, England: Penguin.

SÖRBOM, D. 1975. "Detection of Correlated Errors in Longitudinal Data." *British Journal of Mathematical and Statistical Psychology* 28:138–51.

_____. 1986. "Model Modification." Research Report 86-3. Uppsala: University of Uppsala.

WHEATON, B., B. MUTHÉN, D. F. ALWIN, AND G. F. SUMMERS. 1977. "Assessing Reliability and Stability in Panel Models." Pp. 84–136 in *Sociological Methodology 1977*, edited by D. R. Heise. San Francisco: Jossey-Bass.