Practical problems that are frequently encountered in applications of covariance structure analysis are discussed and solutions are suggested. Conceptual, statistical, and practical requirements for structural modeling are reviewed to indicate how basic assumptions might be violated. Problems associated with estimation, results, and model fit are also mentioned. Various issues in each area are raised, and possible solutions are provided to encourage more appropriate and successful applications of structural modeling.

Practical Issues in Structural Modeling

P. M. BENTLER
CHIH-PING CHOU
University of California, Los Angeles

The methodology of theory testing via structural equation models, which is accepted today as a major component of applied multivariate analysis, is historically relatively new, having been developed in a general and widely accessible form only during the past decade (see Bentler, 1986a, for a review). The process has several basic steps that are probably well known: A model containing random vectors and parameters is developed on the basis of substantive theory, the assumptions underlying the model are used to develop the covariance or moment structure implications of the data, the fixed and free parameters of the model as well as any constraints are imposed, and a statistical method, such as maximum likelihood (ML) or generalized least

AUTHORS' NOTE: *Supported in part by grants DA01070 and DA00017 from the National Institute on Drug Abuse.* *Address reprint requests to P. M. Bentler, Department of Psychology, Franz Hall, University of California, Los Angeles, CA 90024-1563*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 16 No. 1, August 1987 78-117

squares (GLS), is used to estimate the unknown parameters based on a nonlinear optimization method, thus permitting the empirical adequacy of the model to be assessed on the basis of the degree of fit of the model to appropriate sample data. In practice, of course, path diagrams are used to make developing and specifying the model relatively easy, and a general computer program, such as LISREL (Jöreskog, 1977; Jöreskog and Sörbom, 1984) or EQS (Bentler, 1985), which permits a wide range of models and estimation methods to be applied to one's particular model and data, is used to generate the parameter estimates and tests of fit.

Many conceptual elaborations are needed to be able to implement smoothly and appropriately the approach to model building and evaluation summarized above. These important details include not only issues surrounding the translation of substantive theory into a form that can be tested by structural modeling, but also technical details such as mathematical and statistical topics relevant to algebraic model structures, concepts in multivariate analysis, the requirements of a particular computerized procedure, and so on. The first goal of this article is to review the basic assumptions that are needed to assure the appropriate use of structural modeling. Some difficulties that occur, and possible solutions, are discussed. The second goal of this article is to address some further issues that arise in practice, and give suggestions on how to handle these problems. We shall assume that the reader has a general familiarity with structural modeling. See Bentler (1987a) or Long (1983) for introductory presentations.

REVIEW OF BASIC ASSUMPTIONS

Although structural models can be quite easy to set up, estimate, and evaluate with modern computer programs, their output should always be viewed with a certain amount of skepticism: there are many ways in which the methods can fail to reach the lofty goal of evaluating a causal hypothesis. We shall discuss three types of difficulties: conceptual, statistical, and practical.

CONCEPTUAL REQUIREMENTS

It is quite easy to get carried away with the beautiful simplicity with which path diagrams can capture a theory, and with the awesome stacks of computer printouts that epitomize alternative theory-guided views of one's data, thereby losing sight of the fundamental issue of whether some basic conditions for structural modeling have been met. Even the best possible model fit may not protect one from meaningless results.

Obviously, any model will be tested against data obtained from some sample of subjects. An important question that should be answered affirmatively prior to engaging in structural modeling is whether the sample at hand comes from a population that is relevant to the theoretical ideas being evaluated. If the sample cannot be defended as coming from a relevant population, any obtained results may be uninformative about a theory. Thus it is important to know whether the theory one is evaluating should hold for males as well as females, only for a given ethnic group, or only with adolescents. For example, theories of cognitive growth may be relevant only to children or adolescents, but not to young adults. An example of a questionable choice of sample is given in the study of models of female orgasm by Bentler and Peeler (1979). A surprisingly large part of their sample had never had sexual intercourse, and thus did not have a good experiential basis for responding about orgasmic feelings. Luckily, their key findings were validated in a more mature sample (Newcomb and Bentler, 1983).

A closely related issue, easily forgotten in the details of the modeling process itself, is whether the data that might be obtained are gathered under appropriate conditions of measurement (appropriate in relation to the theory under investigation). In the case of opinion surveys, for example, confidentiality of responses may be a requisite to having valid responses to sensitive questions. When evaluating physiological functioning, certain conditions may be called for, such as obtaining urine samples after an appropriate fasting period. When evaluating intelligence in children, appropriately standardized conditions of testing have been recognized as important for many years. However, the fairness of a particular type of test to the ethnic sample being tested, for example, to youngsters who are not native English speakers, may be a question to ponder.

Another important issue involves whether a structural theory is attempting to describe cause-effect sequences that occur over time. The lag required for an antecedent variable to have an effect must be considered. For example, in a model designed to assess the maximum effect of aspirin on relief of pressure headaches, if an antecedent variable is the taking of aspirin, and the consequent variable is relief from headache pain, a contemporaneous measure of relief from pain must be obtained at the causal lag that is appropriate to the maximum expected effect of aspirin: An immediate measure, or one delayed by 24 hours, is likely to show no effect even if one exists at, say, 30 minutes. Whether "instantaneous" causation, with simultaneous mutual influences of variables on each other, makes sense in a model will depend on one's philosophy (e.g., Strotz and Wold, 1960) as well as statistical considerations (e.g., Bentler and Freeman, 1983). A related general point is that longitudinal, or panel, studies may be required to evaluate certain causal sequences. Although some exogenous variables (such as age, sex, race, and so on) may be able to be measured at any time point and still be appropriately incorporated into a model that assumes these variables occur "prior" to others, as causal sequence may not be unambiguously able to be established in many models without incorporating across-time measurement (e.g., Baumrind, 1983). Gollob and Reichardt (1987) have gone so far as to state that cross-sectional models, based on data at one time point only, will virtually always yield biased results on causal effects that are presumed to operate across time. In essence, they argue that a prior measure of the consequent variable is necessary in order to interpret effects of other variables on this consequent variable. A related question is whether a more dynamic, differential-equation type of model is more appropriate to the concepts at hand than the standard model (Arminger, 1986).

An additional conceptual requirement for valid evaluation of theories via structural modeling lies in having theoretically appropriate operationalizations of variables. A given task designed to measure intelligence may not, for example, be appropriate as a variable in a model designed to evaluate Piaget's theory of intellectual functioning. His theory requires specialized assessments, perhaps of such constructs as conservation (Goldschmid and Bentler, 1968), in which equal quantities of water are poured into a narrow and a wide glass, and children are asked whether the narrow and wide glasses have the same amount of water in them (they look unequal).

A final requirement relates to the previous issue, particularly in relation to the use of measured variables as indicators of latent variables. Does a latent variable make sense in a given domain—whether or not a latent variable model fits statistically? In general, a latent variable makes sense when its indicators are logical (theoretical) consequences of the latent variable, not causally related to each other, and correlated sufficiently highly to suggest a common core concept. For example, education, income, and housing quality are from one perspective indicators of a single latent construct. That is, it is easy to conceive of a dimension of social class that has the "haves" at one end, and the "have-nots" at the other end. Increases in the cost of gold, increases in the cost of housing, and increases in the cost of goods over a particular time span would seem to be obvious candidates to be indicators of a latent construct of inflation. However, theory must support the existence of a latent variable for such a construct to make sense in a given model. Thus for some purposes education and income may best not be considered as indicators of social class, since education may be causally related to income. And then, if the construct is meaningful, the particular variables chosen as operationalizations of the construct may be lousy, casting doubt on whether the meaning of the construct has been captured in the operationalizations. Thus a verbal report measure of perceived pressure on the job may, or may not, adequately capture a key idea in a particular theory of stress. And even if it does, the way the respondent's reply is coded into a score may destroy its meaningfulness. In addition, it may well be that the latent space is properly thought of as multidimensional, and a rationale for choosing only a given aspect of such a space may be needed in particular applications. For example, the concept of a single common factor latent variable was introduced by Spearman (1904) in his theory of general intelligence, but recent theory emphasizes a multidimensional construct. Thus performance on college or graduate school entrance examinations will usually contain measures of quantitative as well as verbal skills. The use of a single summary score, the total across both measures, may be justified primarily for practical predictive purposes (see Thorndike, 1985). However, in the prediction of engineering skills, the quantitative score may be more critical. And in developing a structural model with latent variables, a wide enough range of indicators of quantitative skills—say, geometric reasoning, algebraic manipulation, and computational skills (depending on one's theory)—would be desired to assure that the broad construct is well represented. Having an adequate number of indicators also minimizes computational problems (e.g., Anderson and Gerbing, 1984; Boomsma, 1983, 1985).

STATISTICAL REQUIREMENTS

In addition to conceptual requirements associated with structural modeling, there are technical conditions that must be met for the results to be meaningful. Violation of these conditions may make the statistics involved, such as chi-square tests or standard errors, be of questionable quality or possibly downright misleading.

Independence of observations. Current statistical theory used in structural modeling is based on the assumption that data have been gathered from independent observations (cases, subjects, sampling units). It is assumed that responses given by one person will not in any way influence the responses given by another person. In many surveys or telephone interviews, this condition is easy to meet. However, it may also be easy to violate: For example, one may obtain data from a single subject across time, with "observations" referring to repeated measures on the individual. In that case, serial correlation among the responses are quite likely. For example, a mood state that may influence a response at a given time may also influence the response 15 minutes later; however, dependence due to mood fluctuation may not be a problem if the measures are a week apart. Lack of independence may also crop up in innocuous situations: scores from twins may be analyzed, or data may be taken from best friends. It has been argued (e.g., Freedman, 1985) that such a data-gathering design makes the statistical results doubtful since the basic assumption of independence may be violated. Currently, except for specialized regression models, no methods exist for appropriately taking such dependence into account, or of evaluating the assumption by statistical means. Logical arguments must be used.

Identical distributions. The basic theory of structural modeling holds that the same process that describes influences of variables on each other is operating in each and every individual observation or case. (Typically, this condition is described simultaneously with the previous one, under the heading of "i.i.d." assumptions: independent and identically distributed observations.) Stated differently, it is assumed that the path diagram accurately reflects a process that is homogenous across all observations. If such an assumption is false, it is likely that other assumptions—such as a normality assumption—may be violated, which, in turn, may show up in a model not fitting the data. Deviations from a normality assumption can be tested by several means, including Mardia's (1970) test based on multivariate kurtosis that is available in the EQS program. EQS also calculates case contributions to Mardia's test, and these contributions can be used to locate outlying individuals regardless of the distribution. A few extreme outliers are unlikely to be described by a structural model that describes all the remaining observations. Outliers can also be detected by other means (e.g., Comrey, 1985).

If it is suspected that the process theory is different for identifiable subpopulations, such as males or females, it would be appropriate to perform the analysis separately in these populations, based, perhaps, on different models. Actually, the homogeneity assumption can be evaluated in part by a multiple-group or multiple-population model (Bentler, Lee, and Weng, 1987; Lee and Tsui, 1982; Sörbom, 1982), in which key features that differentiate models for different groups can be evaluated.

Simple random sampling. Existing methods in structural modeling are based on the assumption that each of the units or cases in the population has an equal probability of being included in the sample to be studied. In particular, the statistics such as standard errors are appropriate estimators of population parameters only under this assumption. The reason that this is so involves the fact that unadjusted means, variances, and covariances are treated as data to be structurally modeled, and these must be consistent estimators of the corresponding population parameters. When more complex sampling designs are used, the usual covariances as inputs to structural modeling are inappropriate, and the sample means, variances, and covariances must be adjusted to estimate appropriately and consistently the population parameters. This can be accomplished by procedures that give cases differential weight in computations. When such adjustments are needed but not implemented, one must be certain that one understands that the results of modeling will generalize to a population similar to that observed in the sample, but not necessarily to the general population. In some circumstances this drawback may not be crucial, since there is no intent to generalize results to a given population. For example, if the population itself cannot be well defined, then drawing a sample from the population is very difficult if not impossible. To illustrate, the population of cocaine users is easy to conceive, but it is virtually impossible to draw a random sample from the population since most users go to great lengths not to be publicly identified due to legal consequences.

The random sample assumption is usually a reasonable one in practice, although certain data bases are obtained using other methods of sampling. For example, some studies oversample subgroups (such as high-risk subgroups, or subgroups of special interest) that might be critical for a given purpose. More typically, nonrandom samples may occur due to relatively uncontrollable conditions such as the almost inevitable volunteer bias. Thus females are frequently more available and/or cooperative as research subjects, as compared to males, leading to differential representation of the sexes unless special efforts are undertaken. When such differential representation of different groups is encountered, it may be desirable to evaluate the extent to which the grouping accounts for variance in the responses. If this percentage is trivial, analyzing all subjects together is not a significant problem. When variables behave quite differently in different groups, it may be necessary to run a multiple-groups structural model to see the similarities and differences of results across groups. Although recent research has been directed toward developing methods to adjust for biased sample selection (e.g., Bowden, 1986), some controversy exists about whether it is possible to adjust for selection bias (e.g., Little, 1985).

Functional form. The structural models emphasized in this review assume that all relations among variables are linear. This assumption must not only be conceptually appropriate to the theoretical questions being addressed, it should be true empirically as well. It is probably a reasonable assumption when the variables are multivariate normally distributed, or are approximately so distributed after some normalizing transformations. The assumption may be less reasonable if the variables are arbitrarily distributed. In contrast to the case of regression, where numerous diagnostics exist to evaluate assumptions of the model (e.g., Chatterjee and Hadi, 1986), structural modeling diagnostics are virtually nonexistent. Regression diagnostics can, of course, be used equation-at-a-time in multiple-equation path or simultaneous equation models. However, these diagnostics cannot be applied to latent variable models since case scores for the latent variables cannot be determined precisely.

One way to approach the validity of the linearity assumption would be to embed strictly linear models in more complete models that permit nonlinearities, for example, polynomial relations among variables, or interactions between latent variables. If the nonlinear components do not add appreciably to model fit, these could be ignored and the linearity assumption could be accepted.

However, while a general theory for such methods has been developed (Bentler, 1983), existing implementations have been quite specialized (e.g., Etezadi-Amoli and McDonald, 1983; Heise, 1986; Kenny and Judd, 1984; Mooijaart and Bentler, 1986a), and no general computer program is available for use. Thus the use of linear relations is currently based largely on implementability, supported more by successful experience with such models than with evidence that linearity is usually well-justified in social research. Two practical reasons that may be cited for not worrying unduly about linearity is that nonlinear models often do not hold up well in new samples (e.g., Wiggins, 1972), and that nonlinear relations may be approximated by more complex linear models.

The place where nonlinearity is most likely to occur in a predictable manner, and where some progress has been made toward developing methods, has been in models employing categorical variables. Categorical variables raise not only the linearity issue, but also a question about the continuous nature of variables. Continuity will, of course, never be observed in sample data, because the largest number of different scores that could be obtained is the number of subjects in the study. But many variables, such as income, can be seen to be continuous, at least in theory. In practice, they may be categorical because only a few levels of a variable are scored: Individuals may be asked whether they earn less than $5,000, between $5,000 and $15,000, and so on. Other variables, on the other hand, such as sex, may be intrinsically categorical. The methods reviewed in this article are based on the assumption of continuity of dependent variables. Independent observed variables, however, may be categorical.

Olsson (1979), Muthen and Kaplan (1985), and others have argued that the use of categorical variables with methods based on the assumption of linearity and continuity yields distortion of results. One approach to rectifying the problem is to nonlinearly map a categorical variable into a latent continuous variable, and then to develop linear structures for the continuous variables. For instance, tetrachoric, polychoric, or polyserial correlations (e.g., Lee, 1985c; Lee and Poon, 1986) can be used to describe the relations between two underlying continuous variables that, in turn, nonlinearly generate the categorical variables. These indexes of association may be used in structural modeling (e.g., Muthen, 1984; Lee, Poon, and Bentler, 1987). Other approaches are also being developed (e.g., Aitkin and Rubin, 1985; Arminger and Kusters, 1986; Bock and Aitkin, 1981; Bye et al., 1985). However, it has not been verified that these theoretically more appropriate methods generally work better in practice. For example, Collins et al. (1986) found that tetrachorics could yield quite misleading results when compared to the use of technically less appropriate ordinary correlation coefficients with binary variables.

We suggest adopting the following practices for the near future. Continuous methods can be used with little worry when a variable has four or more categories, but with three or fewer categories one should probably consider the use of alternative procedures. If the categorization induces marked nonnormality in variables, a distribution-free method of estimation is called for (see next section). Of course, one should recognize that some distortion will occur in a purely linear model as a result of using categorical variables, yielding a degradation in fit. But this disadvantage must be weighed against the difficulties associated with categorical variable methodology, especially, the restricted number of variables that one will be limited to (this is an even worse problem than with distribution-free procedures), and the necessity, with many of today's methods, of having to make the strong assumption of multivariate normality of the latent continuous variables. (Nonparametric approaches are being developed, however; see, for example, Bye et al., 1985.) Thus categorical variable methodology itself has drawbacks, including a lack of software, that may outweigh the drawbacks of its theoretically less appropriate competitors. Of course, in some situations it may also not make any sense to believe that a continuous variable lies behind a categorical variable. In that case other methods, such as log-linear or latent-class methods, are more appropriate. See Clogg and Eliason's (this issue) or Bonett et al. (1985).

Distribution of variables. Within methodologies for continuous variables, a decision must be made about the distributional form of the variables. Distribution-free methods (Browne, 1982, 1984; Chamberlain, 1982; Bentler, 1983; Bentler and Dijkstra, 1985), of course, do not require such a choice, but they become computationally impractical with models having more than 20-30 variables. Furthermore, their statistics tend to be questionable in small samples, say with less than 200 subjects (Harlow, 1985; Tanaka, 1984). Distributions that are not normal, but in which the variables have a symmetric shape but tails that are heavier or lighter as compared to the normal, are called elliptical if the variables have homogeneous shape or kurtosis (see, e.g., Bentler and Berkane, 1985; Berkane and Bentler, 1987a, 1987b). Then the theory of Browne (1982, 1984), Tyler (1983), Bentler (1983), and Bentler and Dijkstra (1985) can be used to correct the normal theory statistics to lead to appropriate test statistics and standard errors. Distribution-free and elliptical estimators are built into the EQS program (Bentler, 1985). Recent results imply that with some specialized models, such as exploratory factor analysis, the observed variables may in fact not need to be normally distributed and yet, normal theory estimators, standard errors of loadings, and the test statistic may remain correct. This occurs provided that the errors are normally distributed (Amemiya, 1985; Browne, 1985; Mooijaart and Bentler, 1986b; Satorra and Bentler, 1986). These results generalize to a wider class of linear structures when factors are normal and errors are independently distributed (Satorra and Bentler, 1986). Unfortunately, diagnostics to evaluate the relevance of these theories to robustness of statistics in particular applications remain to be developed. Bentler et al. (1986) have made a start in this direction by providing a means for evaluating the distribution of latent variables. However, no diagnostic tests are currently computerized. In the meantime, one can be reassured by simulation evidence that indicates that normal theory ML estimators are almost always acceptable even when data are nonnormally distributed (Harlow, 1985; Muthen and Kaplan, 1985; Tanaka and Bentler, 1985). It is the χ^2 and standard errors that become untrustworthy under violation of distributional assumptions. If one utilizes fit indexes (e.g., Bentler and Bonett, 1980) in addition to statistical criteria for evaluating fit, one's conclusions ought to be reliable.

The use of higher-moment data, such as skewness or kurtosis, to be modeled is a new development (Bentler, 1983). Such methods promise more efficient (lower variance) statistical estimators. However, only sporadic applications of such a theory have been made so far (e.g., Kenny and Judd, 1984; Mooijaart, 1985; Mooijaart and Bentler, 1986a; Heise, 1986).

Covariance structures. Current implementations of structural modeling are based on a statistical theory derived from the distribution of sample means and covariances, and not the distribution of sample-standardized variables having unit variance. Thus the practice of substituting correlation for covariance matrices in analysis is only rarely justified, since the associated statistics will usually be inappropriate (e.g., Bentler and Lee, 1983). While methods have been developed for the structural analysis of correlation matrices (Bentler and Lee, 1983; Lee, 1985a), they are not available in current publicly distributed computer programs.

Large sample size. The exact distribution of estimators and test statistics used in structural modeling is not known. The statistical theory is based on "asymptotic" theory, that is, the theory that describes the behavior of statistics as the sample size becomes arbitrarily large (goes to infinity). In practice, samples can be small to moderate in size, and the question arises whether large sample statistical theory is appropriate in such situations. Even this problem has proven to be hard to study analytically or theoretically, and empirical evidence based on studies with artificial data and models—so-called Monte Carlo studies—have had to be used instead. This research is relatively recent (e.g., Anderson and Gerbing, 1984; Bearden et al., 1982; Boomsma, 1983, 1985; Gerbing and Anderson, 1985; Geweke and Singleton, 1980; Harlow, 1985; Muthen and Kaplan, 1985; Tanaka, 1984; Velicer and Fava, 1987) and has involved only a few types of models, sample sizes, and estimators. Definitive recommendations are not available.

An oversimplified guideline that might serve as a rule of thumb regarding the trustworthiness of solutions and parameter estimates is the following. The ratio of sample size to number of free parameters may be able to go as low as 5:1 under normal and elliptical theory, especially when there are many indicators of latent variables and the associated factor loadings are large. Although there is even less experience on which to base a recommendation, a ratio of at least 10:1 may be more appropriate for arbitrary distributions. These ratios need to be larger to obtain trustworthy z-tests on the significance of parameters, and still larger to yield correct model evaluation chi-square probabilities.

It should also be noted that computational problems during optimization are an inverse function of sample size (e.g., Anderson and Gerbing, 1984; Boomsma, 1983; Gerbing and Anderson, 1987; MacCallum, 1986). While estimating a given model in a large sample may pose no problem, the same model estimated in a small sample may yield such problems as inadequate convergence behavior, boundary or Heywood solutions in which parameters go outside of the permissible range (the classical example being negative variance estimates), inability to impose constraints among parameters, and problems with estimation of standard errors (which may become inappropriately large or very small).

Identified model. Although a recent general statistical theory has been developed for models containing parameters that are not "identified" (Shapiro, 1986), essentially all implementations of structural modeling assume that a model has been specified such that, if the model were true, a single set of parameters θ can reproduce the population covariance matrix. That is, $\Sigma = \Sigma(\theta)$. In contrast, an "underidentified" model will have many different sets of parameters that can equally well reproduce the population covariance matrix. Thus $\Sigma = \Sigma(\theta_1) = \Sigma(\theta_2)$, where θ_1 and θ_2 are different vectors. Parameter identification is a very complex topic, but it can help to think of the uniqueness of parameters as synonymous with identification: A model that does not have identified parameters may have many sets of parameters that can equivalently well account for the data. An example of the complexity of identification was recently given by Bollen and Jöreskog (1985), who showed that previous authors were incorrect to conclude that a factor model is identified if it has factor loadings that are unique with respect to rotation of factors. Such a model may still not be identified, that is, have unique parameters.

The issue of identification can be readily understood in the context of setting the metric in models with latent variables. Suppose one has a model that includes equations of the form V1 = .5*F1 + .2*V2 + E1, where V1 and V2 are observable variables and F1 and E1 are hypothetical variables. Although the scale or variance of every measured variable such as V2 is known in the population (in a sample, it may have to be estimated), this is not true of hypothetical constructs such as factors or errors in variables. The previous equation can be equivalently written as V1 = (10 × .5)*(.1F1) + (10 × .2)*(.1V2) + 10(.1E1), where we have multiplied each coefficient by 10 but then compensated by multiplying each variable by .1. Are we permitted to make such a transformation? When we examine V2 first, it becomes apparent that we would be creating a new measured variable V2 whose variance is .1 × .1 = .01 times as large as the variable we actually have (rescaling by a constant has the effect that the variance is changed by the square of the constant). But V2 has a certain variance and the revised equation would imply that it has a different variance: We are not permitted to make such a transformation. Thus measured variables by themselves create no problem of identification, whereas the situation is quite different for constructs. Since F1 and E1 are hypothetical, with no fixed scale, we could never detect if we replaced the constructs by constructs that are .1 as big (have .01 as much variance). This problem is solved by adopting an arbitrary identification condition that would not permit the rescaling we have illustrated. The best single way of identifying the scale of a latent variable is to fix a path from that variable to another variable at, say, 1.0. When a variable is an independent variable, another way is to fix the variance of the variable at some known value. This is usually done in factor analysis, where factors are fixed to have unit variance. So, we might fix the variance of F1 at 1.0: Then we could not rescale F1 since it would then have a different variance that we do not permit. And we might fix the path from E1 to V1 at 1.0: Then we could not change the implicit 1.0 coefficient in the equation either. Forgetting to fix the scale of unmeasured variables is perhaps the single most frequently made error in applications of structural modeling.

The second most frequently made identification error also involves latent variables. In particular, every latent factor that is meant to account for correlations among some indicators must not only have its scale fixed, but in general it must have effects on (paths to) three or more indicators of that factor. Models with one indicator for a factor will never work (except when a factor is synonymous with a measured variable, that is, the factor is not really a factor). Models with only two indicators will usually run into trouble. While there are exceptions to this rule, particularly when the factor also has nonzero covariances with other factors or variables, such situations must be evaluated quite carefully. Residual variables, such as the E1 variable above, do not need multiple indicators.

Parameter identification may require substantially more care than simply fixing the scale of unmeasured variables or having enough indicators of a factor. Some models require careful attention to particular equations, which may be problematic for additional reasons, or to sets of equations, which may permit several rather than a single solution, or to sets of constraints, which may be redundant (they are not permitted to be so). The most widely known problem involves a "nonrecursive" model, in which two (or more) variables are involved in two-way causation, where a variable is not only an antecedent of other variables, but also a consequent of those same variables. Such models are generally underidentified unless there also exist additional variables that influence, or are influenced by, one but not the other of the variables involved in two-way causation.

A phrase that will frequently be found in discussions of identification involves "overidentification." This is almost always a desirable state of affairs, and refers to a situation where there are fewer parameters in the model than data points. The data to be analyzed are usually $p^* = p(p + 1)/2$ variances and covariances of p variables, and any interesting (testable) model will have fewer than p^* free parameters to be estimated. This difference between p* and the number of parameters yields the degrees of freedom associated with the model fit. If a model has more than p* parameters, it will be underidentified and cannot be tested. A model that has exactly p* parameters will usually be "just-identified," meaning that the parameters are simply transformations of the data, and, hence, the model cannot be tested or rejected. Since there may exist several different sets of parameters that are transformations of the data, just-identified models may not be unique (see, e.g., Bollen and Jöreskog, 1985). Furthermore, a model with less than p* parameters, which is thus nominally overidentified, may nonetheless in particular parts of the model be underidentified (e.g., by an absence of a fixed scale for factors) and hence be not routinely testable until the problem is eliminated.

The theory of identification primarily deals with local, rather than global, uniqueness of parameters. That is, there may exist conceptually quite different parameterizations that reproduce the population covariances equally well. For example, in some models, the direction of certain paths can be completely turned around without affecting the goodness of fit of the model. This topic has hardly been studied. Some interesting examples of this phenomenon, along with a set of rules for generating equivalent path models, are given by Steltzl (1986).

Underidentified models will generally yield statistics that are not strictly correct. While the chi-square value may be trustworthy if an optimum function value was attained, the degrees of freedom may well be understated and thus the p-value (probability) of the model may be too low. If an optimum function value (usually, a minimum) was attained, the computed estimates can be relied upon to reproduce Σ appropriately, but only the identified parameters and their standard errors, or identified functions of all parameters, should be interpreted.

Nested model comparisons. There is little agreement on methods for evaluating the relative merits of two models that are not "nested" or hierarchically related (see Leamer, 1978). Thus an evaluation of the statistical necessity of sets of parameters is limited, under current statistical theory, to a comparison of models in which one model is a subset of the other model, for example, some free parameters in the model are set to zero in a second model. Although current practice is based on the chi-square difference test for making model comparisons, two other equally correct methods exist. Buse (1982) provides an introduction to the key ideas involved in the use of three statistics appropriate to such a purpose: Wald, Lagrange Multiplier, and likelihood ratio tests. Bentler and Chou (1986) and Bentler (1986b) develop these theories for structural modeling.

Increasing the constraints in the more general model will result in a decrease in the number of free parameters, an increase in the number of degrees of freedom, and a consequent increase in the goodness-of-fit χ^2 value. The impact of increasing constraints can be investigated through the Wald (1943) test (Bentler and Dijkstra, 1985; Lee, 1985b), in which only the more general model needs to be estimated. The second approach is to release constraints in the more restricted model, thus increasing the number of free parameters and decreasing the degrees of freedom. Adding new parameters will decrease the goodness-of-fit χ^2 test. The statistical theory for this method is known as the Lagrange Multiplier (LM) theory and has been discussed by Aitchison and Silvey (1958), Silvey (1959), and Lee and Bentler (1980). The application of LM theory requires estimation of only the more restricted model. The third approach is the likelihood ratio approach, or its equivalent. In this approach, both the restricted and less-restricted models are estimated, and the significance of the model-differentiating parameters is investigated by a chi-square difference test. Then χ^2 and degrees of freedom are obtained by calculating the difference between the two goodness-of-fit χ^2 tests, as well as their degrees of freedom. This is currently the standard approach to model comparison.

A priori structural hypotheses. The statistical theory used in structural modeling is based on the fundamental premise that the model itself has been specified completely prior to any analysis of data, that is, the model represents an a priori set of hypotheses. Although one may not know the values of the free parameters of the model (and hence may estimate them in a sample), the entire structure (the particular equations, and variances and covariances of independent variables) should be theoretically derived. If the data are examined, and structural hypotheses are formed after such data snooping, the statistical theory may become incorrect because one may then be capitalizing on chance associations in the data. The effects of capitalizing on chance are particularly acute in small samples, as shown by MacCallum (1986). Adding parameters to an incomplete model on the basis of data snooping can lead to accepting an incorrect true model.

A less serious situation occurs when dropping nonsignificant parameters on the basis of the data, as in backward stepping in regression. It appears that the Wald test for dropping parameters will be more robust than the Lagrange Multiplier test for adding parameters when these tests are data-driven rather than a priori. This is because the Wald test is asymptotically independent of the fit of the more complete model (Steiger et al., 1985). The comparable situation does not occur for the LM test: It depends upon the restricted model.

No parameters on boundary. The statistics of structural modeling assume also that the true parameters are, in the population, in the "interior" of the legitimate parameter space. This assumption is unimportant when dealing with parameters such as regression coefficients that could, theoretically, take on any value. The assumption becomes important when dealing with variances, which must be assumed to be nonnegative. While this is a perfectly natural assumption, estimated variance parameters are sometimes on the boundary (zero) or even in improper regions of the space (negative) and hence not in the interior of the parameter space (e.g., Gerbing and Anderson, 1987). If the population value of the variance is also zero, the model fit tests will be wrong. In that case, a correction to the test statistic as proposed by Shapiro (1985) must be made. However, because it is difficult to know when a population rather than the estimated variance is precisely zero, Shapiro's theory is hard to apply. Thus the researcher must recognize that an assumption is being made, and be prepared to reexamine the assumption when confronted with problematic results.

PRACTICAL REQUIREMENTS

The above conceptual and statistical issues create a number of demands for careful design, data gathering, and analysis in structural modeling. In implementing such requirements, additional practical matters immediately arise. We can review a few of these.

It is easy to become too grandiose when executing a structural model. Most valuable substantive theories are quite complex, and it is easy to hope that most of the complexity can be studied in the context of a single structural model. Rarely is this possible: the data are almost always far more complex than even the best theory, and it is easy to become frustrated in not being able to fit one's model to data. Although one's theory may capture a substantial amount of variation in the data (say, by nonstatistical fit indexes, see Bentler and Bonett, 1980, and Wheaton, this volume), statistical tests can lead to model rejection when only a few effects in the data have been overlooked. In large samples, in particular, even the best model may not fit, since the sample-size multiplier that transforms the fit function into a χ^2 variate will multiply a small lack of fit into a large statistic. To avoid such frustration, without a great deal of knowledge about the variables under study, it is wisest to analyze relatively small data sets, say, 20 variables at most.

On the other hand, it must be recognized that one of the greatest weaknesses in structural modeling lies in excluding key variables that may influence a system. When important control or causal variables are omitted from a model, the parameter estimates of the model will be biased and misleading conclusions can be drawn from an analysis (e.g., Reichardt and Gollob, 1986). Thus one is, in principle, always subject to the criticism of having omitted a key variable. One can only do one's best at ensuring that plausible causal variables are included in a model. But every attempt to include such variables in a model yields a larger model. In turn, this leads to the practical inability to fit models to data noted above. The researcher will have to balance practicality with the ideal of a single, comprehensive model.

Analyzing models with latent variables is an especially demanding problem, since models often do not fit only because a poor measurement structure has been hypothesized for the data. Bentler and Bonett (1980) developed a specification test that evaluates the fundamental adequacy of the measurement model, but it has rarely been used. Ideally, the basic factor structure of the data—how many factors, which variables are good indicators of which factors, the factor intercorrelations—should already be well-known, based on exploratory factor analyses on similar data bases. In that case, generating a complete model that includes a reasonable measurement structure is much easier. Determining the factor structure using structural modeling is a rather difficult and unattractive procedure. It frequently requires so much data snooping to lead one to worry about the quality of any final results.

In the ideal situation, the researcher has in mind not only a single, large structural model, but also a series of submodels that would shed light on key features of the large model. As noted above, nested models can be compared: If the fit of the more restricted model is about as good as that of the more general model, the restrictions can probably be accepted (i.e., the simpler model is chosen and the more complex, rejected). Since there are many possible sequences of submodels that might be entertained, comparing models is an art and requires a good deal of thought.

ISSUES ENCOUNTERED IN PRACTICE

When utilizing structural modeling methods, one will frequently find some problems emanating from the statistical results. These problems may make the results difficult to interpret or even misleading, and hence an awareness of alternative actions that might be taken is valuable. In addition to the problems that are basically created by the estimation procedure, we will also discuss the issues of model improvement.

PROBLEMS IN ESTIMATION

Incomplete data. Although there is no requirement in theory that data used for covariance structure analysis should be complete, that is, that every case should have a score on every variable, current methods were really designed to deal with complete data only. In theory, all that is needed for structural modeling to be appropriate is that the sample covariance matrix S be a consistent estimator of the population matrix and have an asymptotically normal distribution with a known or estimable sampling covariance matrix. Alternatives to the standard matrix can be used. Direct calculation of the sample covariance matrix in fact becomes problematic with missing data, so that programs from general packages such as BMDP may need to be used as preprocessors prior to submitting a job to the current version of EQS. LISREL does have an option for dealing with missing data. Its missing data option may, however, create a problematic sample covariance matrix, as noted below.

A number of recent developments promise better ways of dealing with missing data. For example, Little (1986) has developed a procedure for estimating a sample covariance matrix that is likely to be more robust to outliers than standard methods. Lee (1986) and Van Praag et al. (1985) have provided methods that optimally use all available information during estimation. These methods, however, are not currently available in canned programs.

Covariance matrix not positive definite. Given that a sample covariance matrix to be used in EQS or LISREL exists, one of the first actions of the programs could be to reject the input matrix for analysis. Often this occurs because the input matrix is not positive definite. This can occur for two major reasons. First, there may be linear dependencies among the input variables, in which case the matrix will be singular. Second, the matrix may not be a covariance matrix of real numbers, in which case the matrix will have one or more negative eigenvalues.

Linear dependencies reflect redundancies among variables. Although it may be possible to use an estimation method such as least squares that will accept a singular input matrix, it would generally be desirable to find those variables that are a linear combination of other variables, say, by regression or principal components analysis, and remove them from the input. The reduced matrix then should be acceptable for analysis. Negative eigenvalues tend to occur if the covariance matrix is not computed from raw scores, for example, if tetrachoric correlations are used. They may also occur because a pairwise-delete option was used in generating the covariance matrix (listwise deletion does not have this problem, though it may reduce sample size excessively). There is no generally accepted solution for this difficulty. With missing data, it may be worthwhile to search for outliers in the data, since they may create the problem. Otherwise, it may be necessary to modify the offending entries in the matrix. This can be done by smoothing procedures, for example, by changing negative eigenvalues to a small positive value, or by using special estimators (Theil and Laitinen, 1980). However, the optimality properties of such procedures are not known.

Nonconvergence. This is a common problem that is easy to observe because most computer programs will provide a warning to the user if this has occurred. In covariance structure analysis, parameters are estimated through an iterative process. In other words, the estimates are improved and changed from one iteration to the next until they have stabilized, indicating that they can no longer be improved to obtain a smaller function value. Iterative procedures have a built-in convergence criterion to determine if the change in estimates is so trivial that the iterative process can be terminated. The iterative process can be said to have converged when the change in estimates is smaller than the convergence criterion, and if appropriate derivatives of the fit function are equal to zero. The values updated at the final iteration are the parameter estimates at the solution. However, an infinite or arbitrarily lengthy iterative process may occur if the change in parameter estimates is always large compared to the convergence criterion. This may occur if the model is very nonlinear (e.g., a nonrecursive model), if the model is extremely bad for the data to be modeled (i.e., there are large residuals $s_{ij} - \sigma_{ij}$ even with optimal estimates), if the start values for the parameters are very poor, if unreasonable equality constraints are being imposed, or if critical parameters are underidentified. Corrections for these problems have been mentioned above, or are obvious (e.g., attempt to use good start values, and fix identification problems). The simplest solution, increasing the number of iterations, sometimes yields a convergent solution.

In theory, it may be possible to converge to different solutions, that is, to different local minima rather than a unique global minimum of the function being optimized. In such a case, the model should probably be suspect. However, in spite of the theoretical possibility of local minima, they are very rarely observed. In one case for which multiple solutions were claimed (Rubin and Thayer, 1982), more careful analysis verified that only one minimum actually existed (Bentler and Tanaka, 1983).

Verification that the model specified is the one desired is also a good practice when convergence problems occur. It is easy to make errors that affect estimation, for example, more parameters are fixed at nonzero values than are needed for identification, or inconsistent constraints are imposed. It is valuable to know how the estimates should look at the solution (e.g., small, medium, or large; positive or negative), because a theoretical view of the model can highlight not only problems in job set-up, but also in peculiarities of results. Even LISREL's automatic start values may occasionally yield nonsensical results for nonstandard models, and these can cause nonconvergence. If Σ is singular based on initial estimates, further iteration may not be possible. This can occur because error variances are too small (initially, the variance of an error variable should be as large as possible, but less than that of the corresponding measured variable) or the factor loadings, paths, or variances are too large. In general, the scale of observed variables should be considered when determining the appropriate size of parameters.

Empirical underidentification. Lack of parameter identification will certainly contribute to failures of convergence, but there is no 1:1 relation. Identified models may not converge, and underidentified models may converge. A more difficult problem to spot is that of empirical underidentification, in which a model is actually theoretically identified under most conditions but some special data-related problem occurs that makes the model underidentified (Kenny, 1979; Rindskopf, 1984). This may show up as a nonconvergence problem, as a problem with the information matrix (see below), or a problem with some parameter estimates.

In general, a latent variable model that has only two indicators of a factor will be underidentified since three indicators (another factor may serve as an indicator of the factor as well) are needed at a minimum. So if a factor has three indicators, one may be tempted to conclude that there can be no problem. However, as noted by Kenny (1979) and McDonald and Krane (1979), if one factor loading is identically equal to zero the model will not be identified. Although one may assume that one's estimates will not be zero precisely, in fact some estimates may approach zero and the consequences of underidentification will be felt. These consequences include, for example, "Heywood" cases in which an error variance goes negative (in LISREL) or is held to the boundary (in EQS), and cases in which parameter estimates become very unstable, that is, have very large standard error estimates. Inadequacy of indicators may arise in many guises: It is possible that other paths from a factor, or covariances with the factor, serve the role as a third indicator to identify a factor. For example, a two factor model with two indicators of each factor (and no complex loadings) will be identified as long as the covariance or correlation between the factors is nonzero. But in practice it may approach zero, creating serious difficulties. It is always important to have as many indicators of a factor as possible, subject to practicality, especially indicators having high loadings on their factors (e.g., Gerbing and Anderson, 1985, 1987; Velicer and Fava, 1987). In the context of questionnaire studies, additional indicators can often be created by breaking down a composite measured variable (say, a total score across many items) into a number of subscores. Since such subscores will tend to be less reliable than the composite, this procedure is most beneficial when the subscores are also reliable.

Singular information matrix. The covariance matrix of the parameter estimates is given by the inverse of a particular matrix that is called, in maximum likelihood theory, the information matrix. A similar matrix is used in all methods. Thus, to obtain standard errors, this matrix must be positive definite, and invertible. When this matrix is close to singular, some of the standard error estimates may be meaningless (very small, or arbitrarily large) because they are generated from a numerically unstable matrix. A singular information matrix may also affect iterations, since the updated parameter estimates hinge upon the matrix as well. If the function is appropriately minimized, the χ^2 value and possibly nonunique estimates can be acceptable even if the standard errors are questionable.

Singularity is often caused by dependence among parameter estimates, and EQS and LISREL will flag such dependencies. These dependencies may arise: from model underidentification, in which case the redundant parameters must be removed; from a poor iterative path, in which case different start values must be used; from inappropriateness of one's model, in which case model modifications must be made; or, possibly, from a poorly conditioned set of input variables, in which the input variables must be rescaled to have more similar variances or, more extremely, one or more variables will have to be removed from the model. It may also be caused by sheer numerical problems associated with inadequate computer precision, which may happen to crop up in a special model but not others due to some unknown combination of events. In that case, little can be done by the user.

The above discussion has focused on the case of no equality constraints among parameters. When constraints are also imposed, the covariance matrix of the parameter estimates must of necessity be singular. The information matrix itself need not be invertible, but a modified or augmented information matrix will still need to be invertible. When a problem exists, EQS will give a similar message in this situation, and the same action should be taken to resolve the problem as noted above.

Inability to impose a constraint. When imposing a constraint has the effect that a model will provide a very poor fit to data, it may not prove possible to impose the constraint. EQS will print a message to that effect if this happens. In that case, the constraint must be relaxed and the model reestimated. The iterative process in EQS also needs start values for parameters that meet the constraint. While EQS will adjust the user's start values to meet this condition, in unusual cases this may not be possible and the user will have to make the correction.

PROBLEMS WITH RESULTS

If it appears that an analysis has yielded an appropriate, converged solution, and no condition codes to warn the user about existing difficulties, it is still desirable to analyze the computer output in some detail to determine whether or not more subtle difficulties with the solution can be observed. These subtle problems, if encountered, suggest problems with the model in relation to the data, rather than problems with the estimation method itself.

Improper variances. The most frequently encountered problem involves variances that are estimated as negative or zero. In LISREL, there are no constraints on variances, and they can be estimated as negative. Such estimates are not only meaningless, they are also inappropriate since, for example, true ML or GLS solutions do not allow negative variances. In LISREL, one may accept the solution as is, rerun the job with negative variances set to zero, or reparameterize to yield nonnegativity (Rindskopf, 1983). Setting a negative variance to zero has the effect of changing the degrees of freedom inappropriately and alters the interpretation of the results. In the EQS program, variances cannot go negative unless the user has changed the program's defaults. Thus zero or boundary variances can occur. In both LISREL and EQS the user must evaluate whether such a zero variance estimate causes problems for the conceptual design of the study. A zero error variance may imply that the measured variable is synonymous with a factor, which may or may not make sense. A zero residual in a prediction equation implies that a dependent variable is perfectly explained by its predictors. If results such as these do not make sense, it may be necessary to modify the design of the study, for example, by adding variables to the input data so that more indicators of a factor are created. If it can be assumed that the zero variance observed in a sample represents the population accurately, Shapiro's (1985) theory indicates that the goodness-of-fit chi-square test is not accurate. No alternative or more accurate test value is, unfortunately, currently available in standard computer programs. On the other hand, if the results do make sense and these "Heywood" cases are isolated occurrences, one need not worry too much, especially in large samples. Boundary solutions can be conceived as indicators that the sample size may be too small for an adequate reliance on large sample theory in the given application, since boundary solutions become much less likely with large samples (e.g., Boomsma, 1985; MacCallum, 1986). However, there may be no alternative to accepting the results since a larger sample may not be available.

Improper solutions may also be a clue that a model has been fundamentally misspecified, so it may be worthwhile to evaluate this hypothesis by considering some quite different models, for example, models with a radically different measurement structure. In some cases, improper solutions can arise from outliers in the data, and deleting the offending cases may eliminate the problem (e.g., Bollen, 1987). EQS provides methods for locating and eliminating outliers. If none of these actions solves the problem, we suggest that the model be accepted with the offending estimate held to the boundary. Usually it does not make sense to eliminate the parameter corresponding to a negative or boundary variance. It is true that fit will usually not be degraded significantly by this practice, since negative estimates are usually not significantly different from zero by z-test (Gerbing and Anderson, 1987). Completely aside from the issue of doing data-based model modification, eliminating a boundary parameter may change a model's form into an undesirable one. For example, in a factor analysis model it usually does not make sense to assume that a variable will be perfectly predictable from the factors, and in a predictive equation it seems a priori unlikely that one should be able to predict a given dependent variable perfectly.

Improper correlations. Covariance or correlation parameters may also go outside the legitimate boundary. Correlations, obviously, must lie in the interval of +1 to −1. Covariances, after being transformed into correlations by dividing by standard deviations, must have the same property; EQS forces this on the solution when the variances are fixed, and correlations greater than one do not then occur. Out-of-bounds correlations can be obtained otherwise, as they can in LISREL where there are no constraints at all. Potential problems can be located as follows: EQS prints out a standardized solution in which all covariances have been transformed into correlations, while LISREL transforms only some of its variables so that some hand calculation may be necessary to determine whether any estimated covariances or correlations are outside the legitimate range.

Correlations at the boundary as well as outside the legitimate range imply that two variables are behaving as if they are identical. Even if other features of the solution are adequate, this implies a problem with the model specification. A very high or improper correlation between factors may occur if more than a few variables have loadings on both factors. Adding indicators that are affected by one but not another factor may help. Changing the causes and consequences of the factor may also help.

More generally, the covariance matrix of the independent variables should be positive definite. No program is currently able to impose this feature on the estimates.

Problem path coefficients. In a completely standardized path analysis solution (provided in EQS), path coefficients can be interpreted as standardized regression coefficients (Wright, 1934). Such coefficients should, generally, lie in the interval +1 to –1. When a coefficient in an equation becomes very large, a specification problem should be suspected. Large coefficients may signal linear dependencies among the predictor variables in the equation, for example. It may be necessary to redefine the variables in the equation, possibly by changing the predictor set. If some of the predictors are factors, it may be desirable to alter their indicators. More generally, causes or consequents of the affected variables may need to be changed.

Although there is no reason that the sign of a beta coefficient must be the same sign as the correlation between predictor and criterion variables, differing signs are usually taken as an indication of a "suppressor" effect. For example, when predicting college GPA from high school grades and SAT scores, since all correlations are positive one would hesitate to see a significant negative beta. Such effects can be located by checking the implied correlation. They may arise due to multicollinearities, in which case they will probably be uninterpretable. Suppressor effects are often not only hard to interpret, but in the area where they were first studied, they were hard to replicate (Wiggins, 1972), and the same appears to be true in structural models. Modifications in model structure may make these paradoxical effects disappear.

Statistical discrepancies. Statistical functions and estimators that have the same large sample distribution should yield chi-square tests and estimates that are roughly similar. Thus normal theory ML and GLS methods should yield equivalent results. When large discrepancies are observed, one should be suspicious about the adequacy of the structural model or the distributional assumption. In EQS, several estimators are easily available in the same run and can be compared, and when ML estimation is obtained, a corresponding GLS (called RLS in the program) chi-square statistic is printed out as well. Unfortunately, it is sometimes hard to make changes that eliminate the discrepancies. They should, at least, be reported.

PROBLEMS WITH MODEL FIT

Even when there are no problems with estimation, or unusual features to the results, a specified model may simply not fit sample data. The next step then is to improve the model. In general, there are two ways to do this. One way is through adding constraints and making the model more restricted. The other is through releasing constraints and making the model less restricted, or more general. In either approach, the proper constraints need to be identified. It is essential that the constraints to be added or dropped should be based on theory.

Typical problems. Perhaps the major problem that leads to the need for model modification is lack of a priori knowledge about the measurement structure of the variables. If at all possible, this should be obtained from prior studies with different data. The measurement structure can be wrong for several reasons: an insufficient number of latent variables is hypothesized; a factor is hypothesized for variables that do not correlate well among themselves; or an extremely restricted cluster-type of loading structure is used, in which a factor directly influences only one measured variable. A highly restricted loading structure will usually result in a relatively complex path structure for the latent variables; in contrast, a less restricted loading structure may permit a simpler path structure. A related problem that frequently occurs, not only in measurement models but path models as well, is that predictor independent variables, or criterion dependent variable residuals, are not allowed to covary. A measurement model with a highly restricted loading structure that forces the factors to be uncorrelated will rarely be appropriate for real data. A similar problem occurs when residuals in factors are forced to be uncorrelated. Even if one has strong theory that predicts lack of correlation, one should be immediately prepared to evaluate the theory against data if the strong model does not fit.

Model modification. Bentler and Chou (1986) proposed two statistical methods to obtain information concerning model improvement. When imposing constraints, both the fit function and the degrees of freedom will increase, but it is hoped that the loss of fit is minimal. When releasing constraints, on the other hand, both the fit function, and degrees of freedom will decrease. Here, it is hoped that the gain in fit is maximal. It is hoped, therefore, that the χ^2 can drop significantly with only a slight decrease in degrees of freedom. The two theories for these methods have been mentioned previously. The Wald (W) theory provides a multivariate test for dropping a set of free parameters, and the Lagrange Multiplier (LM) theory yields comparable information for releasing a set of constraints on parameters, or adding free parameters. In both ML or GLS estimation procedures, the Wald and Lagrange Multiplier theories yield statistics with chi-square distributions, but Bentler and Chou also provided nonstatistical equivalents for least-squares estimates. These tests have been implemented in the EQS program (Bentler, 1986b), as described next.

In the W test, only a multivariate test procedure is executed since the univariate W test is the same as the square of the z-test for each parameter estimate at the solution. Both univariate and multivariate LM tests are performed in the EQS program. The univariate LM test is a special case of the multivariate LM test. It offers a χ^2 value for each fixed parameter in a set to be tested multivariately. The concept behind this univariate test is the same as that of model modification indices in LISREL. However, the univariate test usually provides incomplete information. The univariate statistics are obtained under the assumption that there is no relationship between the various constraints. This is seldom the case. A statistically significant LM test for freeing one fixed parameter may not necessarily remain significant in the multivariate test. Therefore, the information generated by the univariate test can be misleading when several fixed parameters need to be freed to get a well fitting model. We strongly recommend the multivariate test for more adequate and efficient model improvement. Of course, in the selection of constrained parameters to be released in the multivariate LM test, a researcher should carefully examine the theoretical basis for each parameter. In addition to the specification of a parameter set, the theoretical importance of each parameter, or parameter group, compared to others, might also be considered.

The EQS program has been developed to allow parameters to be added or dropped in a stepwise process. Free parameters are dropped from the model one at a time in the W test, while fixed parameters are freed one by one in the LM test. Statistically, the W test is designed to drop the least important free parameters in sequence, and the LM test adds parameters in order of multivariate significance. These procedures can be recognized as variants of the backward and forward stepwise approaches in multiple regression analysis. The complete multivariate tests are obtained at the last step.

Several options on parameter set selections and testing procedures have been implemented in EQS, using the Bentler-Weeks (1980) model matrices Φ, γ, and β. Each matrix is composed of several submatrices, depending on the type of variable combinations involved. For example, in the covariance matrix of independent variables Φ we may be interested in correlations between factors, or in correlations between errors (or even in such unusual correlations as between errors in variables and disturbances in equations). Or, we may be interested in specific subparts of the regression coefficient matrices γ (dependent on independent variables) or β (dependent on dependent variables). A parameter, free or fixed, can be from any of these submatrices. Three types of LM testing procedure are available: sequential, simultaneous, and separate. After the set of fixed parameters is specified, these parameter matrices may be grouped in a predetermined order. The sequential procedure will use this order to prioritize parameters in the test. All the parameters in a submatrix (e.g., factor correlations) that are significant (at a previously assigned level) will be included in the test before the next matrix (e.g., correlated errors) can be considered. The significance of a parameter is defined in terms of the increment of χ^2 contributed by that parameter. Since these groups of parameters are tested sequentially, a more significant parameter from a group with lower priority may not get into the test process earlier than a less significant parameter from a higher priority group. The sequential procedure is especially useful when some groups of parameters are theoretically more important than others. In the simultaneous procedure, all parameters are considered for inclusion on an equal basis. In the separate procedure, separate LM tests are provided for each group of parameters. These results must then be combined subjectively by the researcher. These options are not available in the W test, in which all free parameters are automatically included and parameters are dropped on a simultaneous basis, that is, the least statistically significant parameter will be excluded from the model first. The multivariate LM and W tests can be specified to include or exclude parameters in a sequence that is wholly given by the researcher, or in a sequence that is determined by the empirical size of the increment.

From a theoretical point of view, of course, some parameters might need to remain fixed or freed, no matter what a W test or LM test might find. The user can specify which free parameters should not be considered in the W test and which fixed parameters should be excluded from the LM test. One example is that causal paths involving variables ordered in time have only one direction. A path from a variable at time two to a variable at time one would always be undesirable.

Model modification based on theory yields appropriate statistical tests under typical assumptions usually associated with chi-square tests (e.g., Lee and Bentler, 1980; Bentler and Dijkstra, 1985; Lee, 1985b; Satorra, 1986). If empirical model modification is done using a search procedure to locate the best changes, on the other hand, the probability values given for the statistics may be incorrect and the "true" model may not be found (see, e.g., MacCallum, 1986). One way to establish some validity to the resulting model is to compare the final adjusted model with the originally specified model, for example, by correlating parameter estimates for common parameters across solutions, as is done in most research in our laboratory. This does not address the question of whether newly added parameters are inadequate, but can provide some reassurance regarding the stability of the original parameterization across alternate specifications. Thus if the correlation is in the high .9s, the final solution at least contains most of the same information as the initial solution. Thus the initial solution was basically incomplete. On the other hand, if the correlation is lower, the added parameters also destroy the adequacy of the initial parameterization, suggesting greater problems than simple incompleteness of the initial model. Ideally, of course, one would cross-validate any final model (Cliff, 1983; Cudeck and Browne, 1983).

CONCLUSION

This article has provided a summary of various practical issues in structural modeling. These were addressed from the viewpoint of workers engaged in the theory as well as applications of structural modeling. Evidently, we believe the method to be valuable in social research. However, we would be remiss in not pointing out that some individuals are highly skeptical about the value of structural models. Cliff (1983), for example, has a favorable view toward the theory, but feels that the theory is easily misused when researchers forget basic principles of research. Freedman (1985, 1987) is more extreme in his evaluation. He considers the assumptions underlying the method to be inherently implausible, and the entire field to have essentially no value. As might be expected on the basis of our previous discussion, Freedman's arguments have, in our opinion, no solid foundation (Bentler, 1987b).

REFERENCES

AITCHISON, J. and D. C. SILVEY (1958) "Maximum likelihood estimation of parameters subject to restraints." Annals of Mathematical Statistics 29: 813-828.

AITKIN, M. and D. R. RUBIN (1985) "Estimation and hypothesis testing in finite mixture models." J. of the Royal Stat. Society 47: 67-75.

AMEMIYA, Y. (1985) On the Goodness-Of-Fit Tests for Linear Statistical Relationships. Technical Report No. 10, Econometric Workshop. Stanford: Stanford University.

ANDERSON, J. C. and D. W. GERBING (1984) "The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analyses." Psychometrika 49: 155-173.

ARMINGER, G. (1986) "Linear stochastic differential equation models for panel data with unobserved variables." Sociological Methodology, 16: 187-212.

ARMINGER, G. and U. KUSTERS (1986) "Latent trait and correlation models with indicators of mixed measurement level." Manuscript under editorial review.

BAUMRIND, D. (1983) "Specious causal attributions in the social sciences: the reformulated stepping-stone theory of heroin use as exemplar." J. of Personality and Social Psychology 45: 1289-1298.

BEARDEN, W. O., S. SHARMA, and J. E. TEEL (1982) "Sample size effects on chi-square and other statistics used in evaluating causal models." J. of Marketing Research 19: 425-430.

BENTLER, P. M. (1983) "Some contributions to efficient statistics in structural models: specification and estimation of moment structures." Psychometrika 48: 493-517.

BENTLER, P. M. (1985) Theory and Implementation of EQS, a Structural Equations Program. Los Angeles: BMDP Statistical Software.

BENTLER, P. M. (1986a) "Structural modeling and Psychometrika: an historical perspective on growth and achievements." Psychometrika 51: 35-51.

BENTLER, P. M. (1986b) Lagrange Multiplier and Wald Tests for EQS and EQS/PC. Los Angeles: BMDP Statistical Software.

BENTLER, P. M. (1987a) "Drug use and personality in adolescence and young adulthood: structural models with nonnormal variables." Child Development.

BENTLER, P. M. (1987b) "Structural modeling and the scientific method: comments on Freedman's critique." J. of Educ. Statistics 12: 151-157.

BENTLER, P. M. and M. BERKANE (1985) "Developments in the elliptical theory generalization of normal multivariate analysis," pp. 291-295 in Proceedings of the Social Statistics Section, American Statistical Association.

BENTLER, P. M. and D. G. BONETT (1980) "Significance tests and goodness of fit in the analysis of covariance structures." Psych. Bull. 88: 588-606.

BENTLER, P. M. and C. P. CHOU (1986) "Statistics for parameter expansion and contraction in structural models." Presented at the meeting of the American Educational Research Association, San Francisco.

BENTLER, P. M., C. P. CHOU, and S. Y. LEE (1986) "Distributional consequences of linear latent variable structures." Under editorial review.

BENTLER, P. M. and T. DIJKSTRA (1985) "Efficient estimation via linearization in structural models, pp. 9-42 in P. R. Krishnaiah (ed.) Multivariate Analysis VI. Amsterdam: North-Holland.
BENTLER, P. M. and E. H. FREEMAN (1983) "Tests for stability in linear structural equation systems." Psychometrika 48: 143-145.
BENTLER, P. M. and S. Y. LEE (1983) "Covariance structures under polynomial constraints: applications to correlation and alpha-type structural models." J. of Educ. Statistics 8: 207-222, 315-317.
BENTLER, P. M., S. Y. LEE, and J. WENG (1987) "Multiple population covariance structure analysis under arbitrary distribution theory." Communications in Statistics.
BENTLER, P. M. and W. PEELER (1979) "Models of female orgasm." Archives of Sexual Behavior 8: 405-423.
BENTLER, P. M. and J. S. TANAKA (1983) "Problems with Em algorithms for ML factor analysis." Psychometrika 48: 247-251.
BENTLER, P. M. and D. G. WEEKS (1980) "Linear structural equations with latent variables." Psychometrika 45: 289-308.
BERKANE, M. and P. M. BENTLER (1987) "Distribution of kurtoses, with estimators and tests of homogeneity of kurtosis." Statistics and Probability Letters 5: 201-207.
BERKANE, M. and P. M. BENTLER (1987b) "Characterizing parameters of multivariate elliptical distributions." Communication in Statistics-Simulation 16:193-198.
BOCK, R. D. and M. AITKIN (1981) "Marginal maximum likelihood estimation of item parameters: application of an EM algorithm." Psychometrika 46: 443-459.
BOLLEN, K. A. (1987) "Outliers and improper solutions: a confirmatory factor analysis example." Soc. Methods & Research 15:375-384.
BOLLEN, K. A. and K. G. JÖRESKOG (1985) "Uniqueness does not imply identification: a note on confirmatory factor analysis." Soc. Methods & Research 14: 155-163.
BONETT, D. G., J. A. WOODWARD, and P. M. BENTLER (1985) "Some extensions of a linear model for categorical variables." Biometrics 41: 745-750.
BOOMSMA, A. (1983) On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Nonnormality. Amsterdam: Sociometric Research Foundation.
BOOMSMA, A. (1985) "Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation." Psychometrika 50: 229-242.
BOWDEN, R. J. (1986) "Self-selection biases in correlational studies based on questionnaires." Psychometrika 51: 313-325.
BROWNE, M. W. (1982) "Covariance structures,"pp. 72-141 in D. M. Hawkins (ed.) Topics in Applied Multivariate Analysis. London: Cambridge Univ. Press.
BROWNE, M. W. (1984) "Asymptotically distribution free methods for the analysis of covariance structures." British J. of Mathematical and Stat. Psychology 37: 62-83.
BROWNE, M. W. (1985) "Robustness of the likelihood ratio rest of fit of a factor analysis model against nonnormality distributed common factors." Presented at 4th European meeting of the Psychometric Society and Classification Societies, Cambridge, July.
BUSE, A. (1982) "The likelihood ratio, Wald, and Lagrange Multiplier tests: an expository note. Amer. Statistician 36: 153-157.
BYE, B. V., S. J. GALLICCHIO, and J. M. DYKACZ (1985) "Multiple-indicator, multiple-cause models for a single latent variable with ordinal indicators." Soc. Methods & Research 13: 487-509.

CHAMBERLAIN, G. (1982) "Multivariate regression models for panel data." Journal of Econometrics 18:5-46.

CHATTERJEE, S. and A. S. HADI (1986) "Influential observations, high leverage points, and outliers in linear regression." Statistical Science 1: 379-416.

CLIFF, N. (1983) "Some cautions concerning the application of causal modeling methods." Multivariate Behavioral Research 18: 115-126.

COLLINS, L. M., N. CLIFF, D. J. McCORMICK, and J. L. ZATKIN (1986) "Factor recovery in binary data sets: a simulation." Multivariate Behavioral Research 21: 377-391.

COMREY, A. (1985) "A method for removing outliers to improve factor analytic results." Multivariate Behavioral Research 20: 273-282.

CUDECK, R. and M. W. BROWNE (1983) "Cross-validation of covariance structures." Multivariate Behavioral Research 18: 147-167.

ETEZADI-AMOLI, J. and R. P. McDONALD (1983) "A second generation non-linear factor analysis." Psychometrika 48: 315-342.

FREEDMAN, D. A. (1985) "Statistics and the scientific method," pp. 345-366 in W. Mason and S. Fienberg (eds.) Cohort Analysis in Social Research. New York: Springer.

FREEDMAN, D. A. (1987) "As others see us: a case study in path analysis." J. of Educ. Statistics 12:101-128.

GERBING, D. W. and J. C. ANDERSON (1985) "The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis." Multivariate Behavioral Research 20: 255-271.

GERBING, D. W. and J. C. ANDERSON (1987) "Improper solutions in the analysis of covariance structures: their interpretability and a comparison of alternative respecifications." Psychometrika 52:99-111.

GEWEKE, J. F. and K. J. SINGLETON (1980) "Interpreting the likelihood ratio statistic in factor models when the sample size is small." J. of the Amer. Stat. Assn. 75: 133-137.

GOLDSCHMID, M. and P. M. BENTLER (1968) Concept Assessment Kit: Conservation. San Diego, CA: Educational and Industrial Testing Service.

GOLLOB, H. F. and C. S. REICHARDT (1987) "Taking account of time lags in causal models." Child Development 58: 80-92.

HARLOW, L. L. (1985) "Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study." Ph.D. dissertation, University of California, Los Angeles.

HEISE, D. R. (1986) "Estimating nonlinear models: correcting for measurement error." Soc. Methods & Research 14: 447-472.

JÖRESKOG, K. G. (1977) "Structural equation models in the social sciences: specification, estimation and testing," pp. 265-287 in P. R. Krishnaiah (ed.) Applications of Statistics. Amsterdam: North-Holland.

JÖRESKOG, K. G. and D. SÖRBOM (1984) LISREL VI User's Guide. Mooresville, IN: Scientific Software.

KENNY, D. A. (1979) Correlation and Causality. New York: John Wiley.

KENNY, D. A. and C. M. JUDD (1984) "Estimating the nonlinear and interactive effects of latent variables." Psych. Bull. 96: 201-210.

LEAMER, E. E. (1978) Specification Searches: Ad Hoc Inference with Nonexperimental Data. New York: John Wiley.

LEE, S. Y. (1985a) "Analysis of covariance and correlation structures." Computational Statistics & Data Analysis 2: 279-295.

LEE, S. Y. (1985b) "On testing functional constraints in structural equation models." Biometrika 72: 125-131.

LEE, S. Y. (1985c) "Maximum likelihood estimation of polychoric correlation in r×s×t contingency tables." J. of Stat. Computation and Simulation 23: 53-67.

LEE, S. Y. (1986) "Estimation for structural equation models with missing data." Psychometrika 51: 93-99.

LEE, S. Y. and P. M. BENTLER (1980) "Some asymptotic properties of constrained generalized least squares estimation in covariance structure models." South African Stat. J. 14: 121-136.

LEE, S. Y. and W. Y. POON (1986) "Maximum likelihood estimation of polyserial correlations." Psychometrika 51: 113-121.

LEE, S. Y., W. Y. POON and P. M. BENTLER (1987) "Analysis of structural equation models with polytomous variables."

LEE, S. Y. and K. L. TSUI (1982) "Covariance structure analysis in several populations." Psychometrika 47: 297-308.

LITTLE, R.J.A. (1985) "A note about models for selectivity bias." Econometrica 53: 1469-1474.

LITTLE, R.J.A. (1986) "Robust estimation of the mean and covariance matrix from data with missing values." Under editorial review.

LONG, J. S. (1983) Covariance Structure Models: An Introduction to LISREL. Newbury Park, CA: Sage.

MacCALLUM, R. (1986) "Specification searches in covariance structure analyses." Psych. Bull. 100: 107-120.

MARDIA, K. V. (1970) "Measures of multivariate skewness and kurtosis with applications." Biometrika 57: 519-530.

McDONALD, R. P. and W. R. KRANE (1979) "A Monte Carlo study of local identifiability and degrees of freedom in the asymptotic likelihood ratio test." British J. of Mathematical Stat. Psychology 32: 121-132.

MOOIJAART, A. (1985) "Factor analysis for nonnormal variables." Psychometrika 50: 323-342.

MOOIJAART, A. and P. M. BENTLER (1986a) Random polynomial factor analysis. In E. Diday et al. (eds.) Data Analysis and Informatics. Amsterdam: North-Holland.

MOOIJAART, A. and P. M. BENTLER (1986b) "Robustness of M.L. estimates in exploratory factor analysis." Under editorial review.

MUTHEN, B. (1984) "A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators." Psychometrika 49: 115-132.

MUTHEN, B. and D. KAPLAN (1985) "A comparison of some methodologies for the factor analysis of non-normal Likert variables." British J. of Mathematical and Stat. Psychology 38: 171-189.

NEWCOMB, M. D. and P. M. BENTLER (1983) "Dimensions of subjective female orgasmic repsonsiveness." J. of Personality and Social Psychology 44: 862-873.

OLSSON, U. (1979) "On the robustness of factor analysis against crude classification of the observations." Multivariate Behavioral Research 14: 485-500.

REICHARDT, C. S. and H. F. GOLLOB (1986) "Satisfying the constraints of causal modeling," pp. 91-107 in W.M.K. Trochim (ed.) Advances in Quasi-Experimental Design and Analysis. San Francisco: Jossey-Bass.

RINDSKOPF, D. (1983) "Parameterizing inequality constraints on unique variances in linear structural models." Psychometrika 48: 73-83.

RINDSKOPF, D. (1984) "Structural equation models: empirical identification, Heywood cases, and related problems." Soc. Methods & Research 13: 109-119.

RUBIN, D. B. and D. T. THAYER (1982) "EM algorithms for ML factor analysis." Psychometrika 47: 69-76.

SATORRA, A. (1986) "Alternative test criteria in covariance structure analysis: a unified approach." Under editorial review.

SATORRA, A. and P. M. BENTLER (1986) "Some robustness properties of goodness of fit statistics in covariance structure analysis." Proceedings of the Business & Economic Statistics Section, American Statistical Association, 549-554.

SHAPIRO, A. (1985) "Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints." Biometrika 72: 133-144.

SHAPIRO, A. (1986) "Asymptotic theory of overparameterized structural models." J. of Amer. Stat. Assn. 81: 142-149.

SILVEY, S. D. (1959) "The Lagrangian multiplier test." Annals of Mathematical Statistics 30: 389-407.

SÖRBOM, D. (1982) "Structural equation models with structured means," pp. 183-195 in K. G. Joreskog and H. Wold (eds.) Systems Under Indirect Observation. Amsterdam: North-Holland.

SPEARMAN, C. (1904) "'General intelligence" objectively determined and measured." Amer. J. of Psychology 15: 201-292.

STEIGER, J. H., A. SHAPIRO, and M. W. BROWNE (1985) "On the multivariate asymptotic distribution of sequential chi-square statistics." Psychometrika 50: 253-264.

STELZL, I. (1986) "Changing a causal hypothesis without changing the fit: some rules for generating equivalent path models." Multivariate Behavioral Research 21: 309-331.

STROTZ, R. H. and H.O.A. WOLD (1960) "Recursive vs. nonrecursive systems: an attempt at synthesis." Econometrica 33: 1-41.

TANAKA, J. S. (1984) "Some results on the estimation of covariance structure models." Ph.D. thesis, University of California, Los Angeles.

TANAKA, J. S. and P. M. BENTLER (1985) "Quasi-likelihood estimation in asymptotically efficient covariance structure models." 1984 Proceedings of the American Statistical Association, Social Statistics Section, 658-662.

THEIL, H. and K. LAITINEN (1980) "Singular moment matrices in applied econometrics," pp. 629-649 in P. R. Krishnaiah (ed.) Multivariate Analysis V. Amsterdam: North-Holland.

THORNDIKE, R. L. (1985) "The central role of general ability in prediction." Multivariate Behavioral Research 20: 241-254.

TYLER, D. E. (1983) "Robustness and efficiency properties of scatter matrices." Biometrika 70: 411-420.

VAN PRAAG, B.M.S., T. K. DIJKSTRA, and J. VAN VELZEN (1985) "Least squares theory based on general distributional assumptions with an application to the incomplete observations problem." Psychometrika 50: 25-36.

VELICER, W. F. and J. L. FAVA (1987) "An evaluation of the effects of variable sampling on component, image, and factor analysis." Multivariate Behavioral Research 22:193-209.

WALD, A. (1943) "Tests of statistical hypotheses concerning several parameters where the number of observations is large." Transactions of the Amer. Mathematical Society 54: 426-482.

WIGGINS, J. S. (1972) Personality and Prediction: Principles of Personality Assessment. Reading, MA: Addison-Wesley.

WRIGHT, S. (1934) "The method of path coefficients." Annals of Mathematical Statistics 5: 161-215.

P. M. Bentler is Professor of Psychology at the University of Calfornia, Los Angeles. His research deals with theoretical and statistical problems in psychometrics, especially structural equation models, as well as with personality and applied social psychology, especially drug use and abuse.

Chih-Ping Chou is a Research Associate in the Department of Psychology, University of California, Los Angeles. His research interests are statistical and computational problems in covariance structure modeling, multivariate statistical analysis, and measurement.