Misspecifications That Can Result in Path Analysis Structures

Joan Gallini
University of South Carolina

Structural equation modeling provides the behavioral and social science researcher with a sophisticated methodology to investigate causal relationships among variables within a theoretical framework. However, the researcher must guard against specification error that can result from an inadequate theoretical model. The issue of specification error is a concern in any type of causal structure, from the simplest to the most complex. In this article discussion of specification error focuses on path analysis structures, a restricted type of causal model. Such structures imply rather stringent assumptions. Violations of the assumptions are discussed as possible sources of specification error in path analysis models. Alternative models are formulated and empirically estimated to demonstrate the possible consequences of misspecified models. The discussion is extended to address the issue of specification error in LISREL-type models as well.

Recent attention has been given to the methodology of path analysis in the social and behavioral sciences (e.g., Costner, 1969; Duncan, 1975; Goldberger, 1972; Heise, 1975; Werts & Linn, 1970). Path analysis refers to an approach based on regression analysis for estimating effects among a set of variables. Use of the methodology requires the explicit formulation of a set of variables selected from theory with the direction of the effects specified among the variables. The result is a causal or structural equation model, and a system of structural equations algebraically describes the causal structure.[1]

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 7, No. 2, Spring 1983, pp. 125–137

0146-6216/83/020125-13$1.65

From a broader perspective, path analysis models can be described as restricted cases of structural equation models, such as of the LISREL-type (Jöreskog, 1973, 1978). Briefly, a LISREL (linear structural relationship) model includes latent or unobservable variables that are evaluated via a set of multiple indicators or measured variables. Such structures thus account for the unreliability or measurement error associated with each variable. In contrast, the more restricted path analysis models represent all variables under investigation as directly observed and perfectly reliable measures in the model. In addition, use of path analysis imposes somewhat stringent assumptions about the error terms in the structural equations that can otherwise be relaxed in a more general model.

The nature of the structure dictates the appropriate statistical technique to apply to estimate the specified effects. Thus, it is extremely important that the researcher considers the tenability of the

[1]Causal modeling is often used interchangeably with structural equation modeling. However, it should be noted that the "causal" term is frequently used in a descriptive sense. The reader is cautioned against a literal interpretation of "cause," since rigorous causal inference can only be obtained by means of carefully controlled experiments. As Jöreskog (1981) has pointed out, however, sometimes causal inference in the social sciences can be obtained by means of tightly specified models such as those estimated from time series data.

assumptions implied by a given causal structure. A misspecified model can result in biased estimations. Such bias due to model misspecifications has been termed specification error.

A given specification error can have serious consequences in one instance and negligible effects in another. In any event, it is always reasonable to examine a model for possible misspecifications. Often it is useful for the researcher to start with the more restricted version of a structural model formulated from theory. The consequences of possible specification error can then be examined on the estimated effects by entertaining alternative models that relax some of the assumptions.

It is the purpose of this article to identify and discuss four types of specification error in path analysis-type structures. Empirical illustrations are provided to demonstrate different instances in which specification error may or may not affect the estimations.

As a peripheral outcome, the article illustrates a useful approach to investigate rival hypotheses about different theoretical schemes. That is, a hierarchy of alternative models results by relaxing certain assumptions thought to be misspecifications in a given path analysis model. In turn, a comparison among the alternative solutions provides an assessment of the "most reasonable" theoretical structure.

Path Analysis

The methodology of structural equation modeling is based on regression analysis. Thus, path analysis and the more general LISREL-type models are linear, additive models, and generally require interval measures. Billings and Wroten (1978) examined violations of linearity, additivity, and interval measures in path analysis models. Nygreen (1971) addressed interactions in path analysis.

Path analysis models represent a restricted type of causal model. Therefore, an additional set of assumptions (specifications) are implied in path analysis models. Those assumptions are the topic of discussion in this article. First, a brief overview of path analysis is provided, followed by a discussion of the assumptions.

Overview

This section is didactic in nature. It provides a brief overview of the path analysis methodology. For a more detailed account, the reader is referred to Duncan (1966, 1969, 1975), Heise (1975), Land (1969), Pedhazur (1980), and Wright (1934, 1960).

Causal modeling employs a set of techniques for estimating specified effects among variables that have been sequentially ordered prior to any data collection and analysis. The structural or causal scheme is derived from substantive theory. The model specifications are translated into a system of simultaneous equations. Consider, for example, the simple three-variable causal model in Figure 1. The model specifies a direct causal dependency of B on A and of C on B. In addition, A is said to indirectly affect C via variable B.

Variable A is called an exogeneous variable whose causes are not considered in the present model. Variables B and C are endogenous, with B said to be predetermined with respect to C. Relationships among the endogenous variables are evaluated within the system. The e_i's (i = 1,2) represent the errors or disturbances due to unexplained variance at each stage of the model.

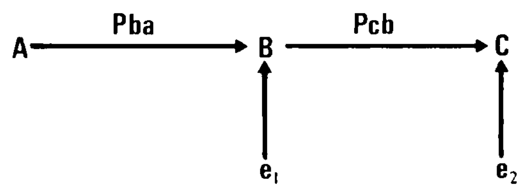
The model is mathematically represented by Equations 1 and 2:

$$B = p_{ba}A + e_1 \qquad [1]$$

$$C = p_{cb}B + e_2 \qquad [2]$$

The p_{ij}'s (i = dependent or effected variable, j = causal or independent variable) are called path or structural coefficients that need to be estimated. Structural coefficients are indices of the magnitude of variable j's influence on the later variable i. For example, p_{ba} represents the direct effect of A on

Figure 1
A Three-Variable Causal Scheme



B. The product $p_{cb}p_{ba}$ represents the indirect influence of A on C. An appropriate statistical technique is used to estimate the model parameters. The technique to be applied depends on the type of causal model.

Basically, there are two types of causal models: (1) models in observable or measured variables and (2) models in unobservable or unmeasured (latent) variables. The causal flow among the variables in causal models can be designated as recursive (unidirectional) or nonrecursive (reciprocal). (For further reading, see, e.g., Bielby & Hauser, 1977; Wolfle, 1980.)

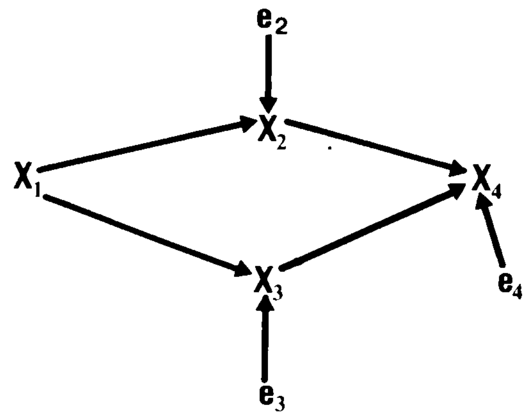
The more general causal models of the LISREL-type include latent or unobservable variables. The latent variables are evaluated by multiple indicators or measured variables. Such models thus account for the unreliabilities of the measurement. The LISREL system can be implemented by the LISREL computer program (Jöreskog & Sörbom, 1978, 1981), which employs the maximum likelihood estimation technique.

On the other hand, path analysis models represent a restricted type of structural equation model in that all the variables are treated as observable measures. Consequently, such models do not allow for the presence of measurement error in the variables. Ordinary least squares (OLS) techniques are generally applied to estimate the specified effects in the typical path analysis structure. Thus, OLS assumptions apply to such models. There are additional restrictions implied by path analysis models. The interpretation of the assumptions in terms of path analysis model specifications are examined below.

Assumptions Implied by Path Analysis Structures

Assumption 1. Residuals are uncorrelated with the predictor variables. In terms of path analysis, this means that the disturbances are uncorrelated with the exogenous variables and all predetermined endogeneous variables. Such an assumption is implied by the specifications of the path analysis model in Figure 2. This means that $E(X_1, e_2) = E(X_1, e_3) = E(X_1, e_4) = 0$, or that the disturbances are uncorrelated with the exogenous variable X_1. In addition, $E(X_2, e_3) = E(X_2, e_4) = E(X_3, e_4) = 0$ is implied by the model specifications. The latter restriction means that the disturbances are uncorrelated with the predetermined endogenous variables.[2]

Figure 2
A Hypothetical Path Analysis Model



Assumption 2. An additional assumption in path analysis is that the disturbances of the endogenous variables are mutually uncorrelated. In model specification terms this means the $E(e_2, e_3) = E(e_2, e_4) = E(e_3, e_4) = 0$.

Assumptions 1 and 2 above imply the following model specifications:
1. Any variable omitted from the model is unrelated to the predetermined variables.
2. The directional flow among the variables is recursive.
3. The exogenous and endogenous variables are measured perfectly.

The omitted variable, recursivity, and no measurement error assumptions place some stringent restrictions on the structure of path analysis models. Path analysis models are thus viewed as nested or special cases of expanded models, such as nonre-

[2]Each dependent variable is predetermined with respect to any other dependent variable that occurs later in a causal order. This means that the predetermined variables in an equation are uncorrelated with the disturbance(s) of that equation in the system (Duncan, 1975).

cursive models in observables or recursive models in latent variables. If the above assumptions are valid, OLS techniques will yield unbiased estimates. However, it is necessary to entertain the plausibility of violated assumptions or a so-called misspecification. A misspecification in a model results in a specification error. Sources of specification error can be identified by examining the parameter estimates of alternative causal models that have relaxed a particular specification (or assumption).[3]

The present article does not attempt to address all possible sources of misspecification in causal models or propose a taxonomy among the specification errors addressed. For example, the theoretical ordering assigned among the variables can be a specification error in any type of causal structure. Instead, it is the intent of the present discussion to investigate four types of specification errors characteristic of path analysis models, since they are the most restricted type of causal structures.

Sources of Specification Error

Specification Error 1: Uncorrelation between Disturbances and Exogenous Variables

A major assumption of path analysis is that a lack of correlation exists between the disturbance of an endogenous variable and exogenous variables. The assumption implies that, basically, no exogenous variable is omitted from the initial model that is an "important" correlate or causal variable of other variables in the system. If the assumption is not valid, OLS procedures will yield biased estimates of the path coefficients. The bias is the result of a specification error due to an omitted variable.

Consider Path Model A in Figure 3. The path model is represented by the set of Equations 3 through 5.

$$X_2 = p_{21}X_1 + e_2 \qquad [3]$$

$$X_3 = p_{32}X_2 + p_{31}X_1 + e_3 \qquad [4]$$

$$X_4 = p_{43}X_3 + p_{42}X_2 + p_{41}X_1 + e_4 \qquad [5]$$

Exogenous variable X_1 = a general aptitude measure, and endogenous variables X_2, X_3, and X_4 represent achievement measures collected at different points in time.[4] Briefly, the specified relationships describe a direct dependency of initial learning (X_2) on an individual's aptitude (X_1). In addition are the direct causal influences of previous learning of objectives on later learning (e.g., X_2 on X_3) during the instructional sequence.

In this model the assumption is made that $E(X_1, e_2) = E(X_1, e_3) = (X_1, e_4) = 0$. In addition it is assumed that $E(e_2, e_3) = E(e_3, e_4) = E(e_2, e_4) = 0$. OLS methods can be used to estimate the path coefficients, given that the present specifications above about disturbances are valid. Suppose, however, that an important causal variable of X_2 and X_3 is omitted from the model. If so, OLS techniques will be inappropriate to estimate the specified effects of Model A.

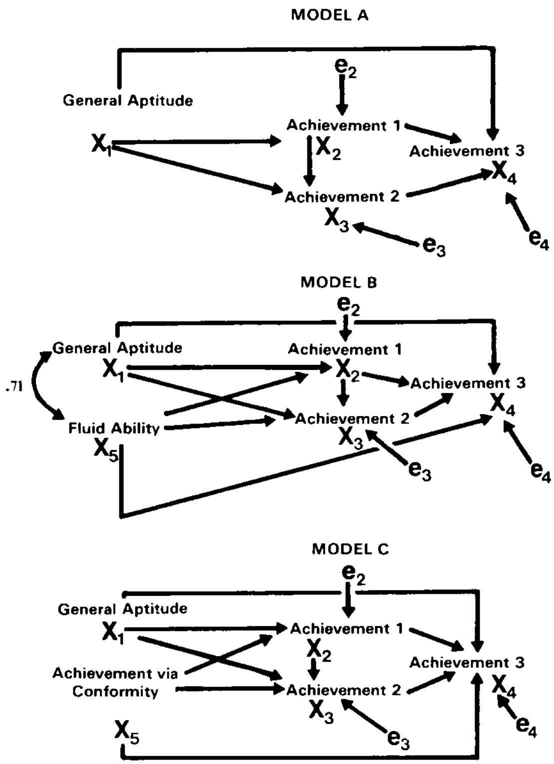
One way of addressing possibilities of an omitted variable is by introducing the variable into the model. For example, the related theory underlying Model A specifications suggests the importance of two aptitude complexes important for learning (Cronbach & Snow, 1977). The cognitive complex consists of crystallized intelligence, fluid ability, and visual ability (GcGfGv). The affective component consists of achievement via conformity (Ac), achievement via independence (Ai), and general anxiety (Ax).

In Model A, exogenous variable X_1 represents only the Gc factor of the aptitude complex. Omis-

[3]In addition, statistical techniques are available for testing the overall fit of an entire causal model. One approach in path analysis is to reproduce the correlation matrix. A "good" fit is reflected when the reproduced correlations do not exceed .05 (see Kerlinger & Pedhazur, 1973). Goodman (1973) has described some statistical techniques for examining whether a model fits the data. Further technical discussions are provided by Bentler and Bonett (1980).

[4]Selected variables were abstracted from the study "A Path-Analytic Study of Aptitude-Learning Relations Over Time" (Burns & Gallini, 1981).

Figure 3
Path Analysis Models of Aptitude-Learning Outcome Relationships Over Time



sion of the Gf and Gv factors, if indeed causally important to the other variables in the model, will result in a nonzero correlation between X_1 and e_2, e_3, and e_4. This, of course, would be a violation of a basic path analysis assumption.

One way to investigate the specification error is to introduce the omitted variable and to observe what happens to the estimates. For illustration, the Gf factor has been introduced into the original model. The revised Model B now shows an additional variable and a direct causal link from Gf to each achievement measure. Table 1 presents the OLS estimates obtained under this model (column 2) and those from the four-variable Model A (column 1). Note the changes in the causal coefficients when the Gf variable is entered. For example, the path p_{21} from general aptitude (X_1) to achievement 1 (X_2) decreases from .65 to .41. The difference can

Table 1
Path Coefficients Obtained Under the Four-Variable Model and the Two Alternative Five-Variable Models Illustrated in Figure 3

Path Coefficient	Model A	Model B	Model C
p_{21}	.65	.41	.52
p_{25}	-	.33	.01
p_{31}	.50	.43	.52
p_{32}	.20	.17	.16
p_{35}	-	.15	.08
p_{41}	.41	.49	.46
p_{42}	.29	.35	.26
p_{43}	.07	.11	.06
p_{45}	-	-.21	-.09

possibly be explained by the specification error due to a correlation of disturbance e_2 with exogenous variable X_1 in Model A, where the Gf factor was omitted.

It is useful to consider the conditions under which an omitted variable will lead to serious misinterpretations. In the example, the omitted Gf variable showed a somewhat high correlation of .71 with the exogenous variable Gf. In addition, it showed a significant causal effect (p_{25} = .33) on achievement 1. Thus, an omitted variable that is a strong correlate and cause of other variables in the model will lead to serious bias (Kenny, 1979). To further illustrate this point, consider an omitted variable that is not a strong causal variable. Model C of Figure 3 considers such a third variable, X_5, achievement via conformity (Ac). The Ac is a factor of the affective complex suggested by Snow (1976). Thus, while the underlying theory suggests its importance, the present data set showed a low correlation of .25 of Ac with the Gc exogenous variable. In addition, its causal effects on other endogenous variables in the model are low. Thus, the consequence of omitting the Ac variable from the model are trivial—at least from the perspective of drawing erroneous causal inferences. This is reflected by a comparison of estimates in columns 1 and 3 of Table 1. The comparison shows relatively small changes between the path coefficients from the original model to Model C.

The value, then, might be questioned of maintaining a variable such as Ac in the model. For one, the underlying theory does reference its importance in an instructional learning sequence. Its importance may become more evident in other data sets. Even more important, introduction of the variable adds increased validity to the original path model by the specifications of some zero causal paths (Kenny, 1979).

Undoubtedly, the test of the uncorrelated disturbance assumption can become a cumbersome task, especially if there are several omitted variables under consideration. As Duncan (1975) has pointed out, this is the "hard part" of structural equation modeling. The calculations are the "easy part." There are, however, different strategies for reducing the many possibilities. One way can be through experimental design. That is, the researcher can control outside variables. For example, a selection of subjects from only one stratum of a particular variable will eliminate the effect of that variable (Billings & Wroten, 1978).

An examination of the residual scatterplots can be useful (Draper & Smith, 1966). Deviations from a horizontal band suggest that the disturbances are correlated, which might be due to an omitted variable. The plots should be re-examined after one or two primary outside variables are included.

Even more useful is that the researcher start with a sound theory and use the theory to carefully select and to explicitly order the variables in a causal framework. The empirical illustrations above demonstrate the need to include important causal variables in the system. At the same time, the model is improved by specifying zero paths among variables.

Specification Error 2: Zero Correlations between the Disturbances

The assumption of mutually uncorrelated disturbances of the endogenous variables is a path analysis assumption. Billings and Wroten (1978) conceptually discuss the consequences of violating the assumption. It would be useful to empirically demonstrate the consequences of such a misspecification.

Consider the initially proposed Model A of Figure 4 with variables X_1 = general aptitude, X_2 = number of hours accrued in advanced college credits, X_3 = transfer or nontransfer student, X_4 = undergraduate GPA, and X_5 = years in undergraduate major program.[5] The model is represented by Equations 6 and 7.

$$X_4 = p_{41}X_1 + u \qquad [6]$$

$$X_5 = p_{52}X_2 + p_{53}X_3 + v \qquad [7]$$

The model specifies that $r_{uv} = 0$ as well as uncorrelated exogenous variables and disturbances. Thus, assuming the above specifications are tenable, OLS will yield unbiased estimates for the path coefficients.

Suppose, however, that the related theory suggests that another variable (not considered in the present model) is a common causal variable of X_4 and X_5. Thus, exclusion of that variable means that the $r_{uv} = 0$ will be a misspecification (as well as the specified zero correlations between exogenous variables and disturbances). Under such conditions this means that the assumption about uncorrelated disturbances is violated. An alternative model that allows the disturbances to correlate is appropriate to consider to deal with the specification error.[6] Model B of Figure 4 specifies correlated disturbances. Econometricians refer to this special case of a path analysis model where errors are correlated as a seemingly unrelated system (Pindyck & Rub-

[5]The selected variables are drawn from a larger data set of the study "Evidence of an Adaptive Level Grading Practice through a Causal Approach" (Gallini, 1982.)

[6]It should be noted, however, that an alternative model may not always yield estimates that are all that different. Such findings suggest that the alternative structure will not improve the fit of the model to the data.

infeld, 1981). That is, the equations in such a system are linked by the correlation of errors across the equations, and the errors are said to be serially correlated.

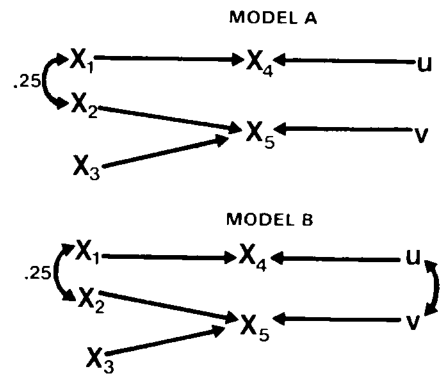
The path coefficients of each equation in a seemingly unrelated system can ordinarily be estimated with OLS. The OLS estimates will be consistent and unbiased values. However, the efficiency of OLS estimates in seemingly unrelated systems is affected. The estimates can be improved by explicitly accounting for the correlation specified among the errors. A method called Zellner estimation (1962) can be used. Briefly, generalized least squares (GLS) is applied to estimate the entire system of equations. The objective of GLS is to find parameter estimates in the most efficient manner. In doing so, the linear model is generalized to include both serial correlations and heteroscedasticity in the variance-covariance matrix of the error terms.[7] (For further reading on the GLS technique see Browne, 1974; Lee & Bentler, 1980; Pindyck & Rubinfeld, 1981; Zellner, 1962; Zellner & Theil, 1962.)

Table 2 shows the estimates obtained under OLS and GLS. In the present data set there are small differences between the estimates under the two solutions. This finding implies that the $r_{uv} = 0$ specification is not a serious error in Model A. In fact, the OLS estimates would be appropriate in such a case. However, to the extent that the solutions yield larger discrepancies, the GLS estimates should be used since they are more efficient. That is, the standard errors associated with the GLS estimates will be relatively smaller than the OLS, since the GLS procedures explicitly account for the correlation between the disturbances.

[7]A clarification is made concerning GLS estimations. Time series studies typify causal structures with serially correlated errors. That is, disturbances associated with observations at one time period carry over into future periods. Ordinarily, time series or longitudinal studies concentrate on cases where the focus of the data matrix are correlated. The Browne-type GLS deals with another type of serial correlation, i.e., column dependency among the estimates of the variance-covariance matrix.

Figure 4

Path Analysis Models of Specified Relationships with GPA in Undergraduate Program Majors, with X_1 = Aptitude, X_2 = Advanced College Credit, X_3 = Student Status, X_4 = GPA, and X_5 = Length of Program



Specification Error 3: Recursive Relationships among Variables

Use of path analysis requires the assumption of recursivity. That is, it is assumed that the underlying theory specifies a unidirectional causal flow among the variables. The exemplary models thus far have assumed recursive relationships.

Suppose that recursivity is a misspecification in a model. Hypothetically speaking, the related theory might allude to a dependency of Y on Z but at the same time of Z on Y. Thus, both variables are causal and effected variables in the system, which economists refer to as simultaneous equation systems. The OLS estimates will be biased in simultaneous systems.

Reciprocal or nonrecursive relationships can be evaluated within a causal framework. The aptitude data set will be used to empirically illustrate the specification error and an alternative strategy to OLS.

Model A of Figure 5 represents a five-variable recursive model. Exogenous variables are X_1 = general aptitude measure, X_2 = fluid ability measure (i.e., Raven Progressive Matrices; Raven, 1958), X_3 = general anxiety measure. The endogenous variables are X_4 = achievement via independence and X_5 = achievement via conformity.

Table 2
Path Coefficients and Their Standard Errors (S.E.) Obtained Under Ordinary Least-Squares (OLS) and Generalized Least-Squares (GLS) Procedures for the Models in Figure 4

Path Coefficient	Model A OLS Estimate	Model A S.E.	Model B GLS Estimate	Model B S.E.
p_{41}	.002	.00	.001	.00
p_{52}	−.120	.11	−.100	.08
p_{53}	.820	.82	.840	.78

As it stands, Model A rules out a reciprocal relationship between Ac and Ai. The only relationship specified between the two endogenous variables is that Ai affects Ac ($p_{54} \neq 0$), but $p_{45} = 0$.

The mathematical form for Model A is shown by Equations 8 and 9.

$$X_4 = p_{41}X_1 + p_{42}X_2 + p_{4u}u \quad [8]$$

$$X_5 = p_{53}X_3 + p_{54}X_4 + p_{5v}v \quad [9]$$

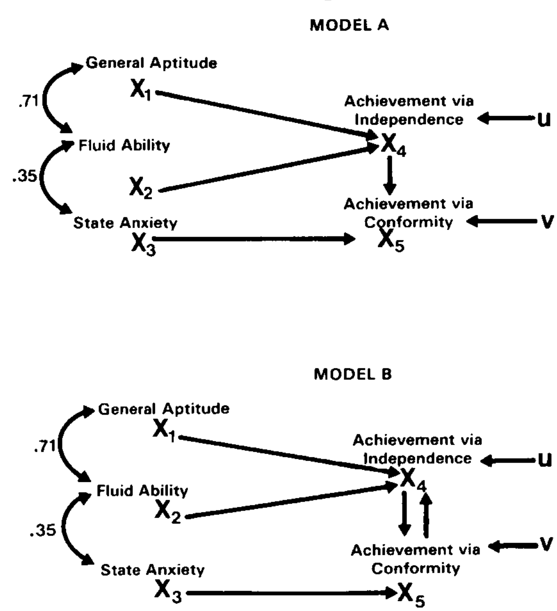
In this type of structure it is assumed that the disturbance terms in each equation above conforms to the least squares classical assumption that $u \sim N(0, \sigma_2^2)$ and $v \sim N(0, \sigma_v^2)$, so that $E(X_1, u) = E(X_1, v) = E(X_2, u) = E(X_2, v) = E(X_3, u) = E(X_3, v) = 0$. In addition, $E(X_4, v) = 0$ and consequently $E(u, v) = 0$.

However, suppose the $p_{45} = 0$ is a misspecification. This is a plausible source of specification error in Model A. Substantively speaking, the Ac scale of the *California Psychological Inventory* (Gough, 1975) reflects "those factors of interest and motivation which facilitate achievement in any setting where conformance is a positive behavior (p. 11)." The Ai reflects "those that are not empirically independent nor bipolar representations of the same construct." In fact, the manual reports empirical correlations of .38 and .39 for males and females, respectively, between the two scales. Thus, it is reasonable to consider nonzero paths from Ai to Ac as well as from Ac to Ai.

Alternative Model B portrays a nonrecursive relationship between Ai and Ac. Equations 10 and 11 algebraically describe the model.

$$X_4 = p_{41}X_1 + p_{42}X_2 + p_{45}X_5 + p_{4u}u \quad [10]$$

Figure 5
Path Analysis Models of Aptitude-Learning Outcome Relationships Over Time



$$X_5 = p_{53}X_3 + p_{54}X_4 + p_{5v}v \qquad [11]$$

With the present specifications, OLS estimates will be biased. A logical explanation shows why this is so. In estimating Equation 10, OLS will be appropriate only if X_5 is uncorrelated with u. A change in u will produce a change in X_4; but because X_5 is a function of X_4 (Equation 11), the change will effect a change in X_5 as well. This means that X_5 is correlated with u, and so OLS will yield biased estimates. The same logic explains the correlation between X_4 and v.

Two-stage least squares (2SLS; Theil, 1971) can generally be used to obtain unbiased estimates in such systems. Conceptually speaking, the first stage of the 2SLS procedure involves the creation of a so-called ''instrumental variable'' or new variable from the original specifications. These new variables are free of confounding effects from correlated disturbances. They are then employed in a second regression generation analysis to estimate the structural coefficients (Heise, 1975).

Table 3 presents the path coefficients obtained from three different solutions. Column 1 shows the OLS estimates obtained under Model A specifications, where recursivity was assumed. Columns 2 and 3 respectively provide the OLS and 2SLS estimates obtained under Model B that rules out recursivity (i.e., it considers the rival hypothesis of nonrecursivity). The OLS estimates for p_{45} and p_{54} in column 2 differ considerably from the 2SLS estimates in column 3. It is probably safe in concluding that the nonrecursive structure would yield an improved fit to the data.

Specification Error 4: Lack of Measurement Error in the Variables

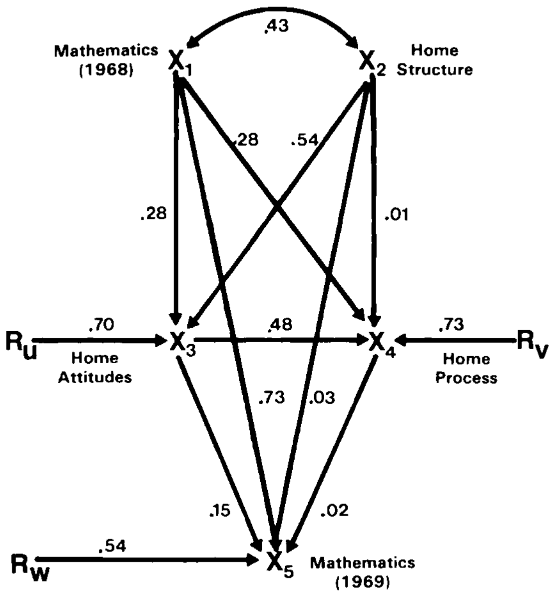
Another restriction of path analysis models is that no allowance is made for measurement error in the variables. The assumption refers to an ''error in variable'' problem due to an unreliability in measuring the variables. If the absence of measurement error is a misspecification, this will lead to attenuation bias in the path coefficients.

There are several procedures that can be used to correct for attenuation bias (e.g., Kenny, 1979). Operating within a causal modeling framework, one such way introduces latent or unobservable variables in the model. Each latent variable is evaluated by one or more observable indicators or measured variables. When only one indicator is used, the model basically represents the special case of a path analysis model. However, considerable measurement error is allowed to enter path analysis models. To overcome the restriction, a combination of multiple indicators for the latent variables can be selected to yield a better measurement of the latent variables within the LISREL system of structural equation models.

Table 3
Path Coefficients and Their Standard Errors (S.E.) Obtained for the Two Models of Figure 5, A Recursive Model (Model A) Under Ordinary Least-Squares (OLS) Techniques and an Alternative Nonrecursive Model (Model B) Using Ordinary Least-Squares and Two Stage Least-Squares (2SLS) Estimation Techniques

	Model A		Model B			
Path Coefficient	OLS Estimate	S.E.	OLS Estimate	S.E.	2SLS Estimate	S.E.
p_{41}	.30	.004	.11	.110	−.27	.100
p_{42}	.16	.120	.23	.004	.11	.080
p_{45}	–	–	.40	.080	.12	.010
p_{53}	.53	.050	.46	.050	.52	.050
p_{54}	.35	.110	.35	.110	.62	.080

Figure 6
Cooley-Lohnes Path Model for Home Environment and Mathematics Achievement (Adapted from Cooley and Lohnes, 1976, p. 158, with Permission)



The LISREL system of causal modeling is described by a structural and a measurement component. The structural model describes the linear relationships among the constructs or unobservable variables, and the measurement component characterizes the regressions of the indicators on the constructs. The specified effects can be assessed by using the LISREL IV or V computer program. The obtained solution yields estimates of measurement error through maximum likelihood techniques. (For further technical details about applications of LISREL see, e.g., Bentler, 1980; Jöreskog & Sörbom, 1979; Lomax, 1982; Maruyama & McGarvey, 1980.)

An example from Sörbom and Jöreskog (1981) is used to illustrate consequences of the no-measurement-error constraint in path analysis models.[8]

[8]Portions of the Sörbom and Jöreskog (1981) paper have been used here to facilitate the present discussion of specification error. For further details of their analysis, the reader is referred to the original paper.

Their empirical example involves a reanalysis of Keeves' (1972) data that reflected the importance of the home environment for school achievement in a path analysis model (Cooley & Lohnes, 1976). The path analysis model extracted from Cooley and Lohnes (1976, chap. 4) appears in Figure 6. In the path diagram X_1 = initial mathematics ability, X_2 = the structural dimension of the home, X_3 = the attitudinal dimension of the home, X_4 = the process dimension of the home, and X_5 = final mathematics achievement. The three home variables were generated from a principal components solution. The OLS regression showed that neither the structural variable nor the home processes significantly influenced final achievement (.03 and .02).

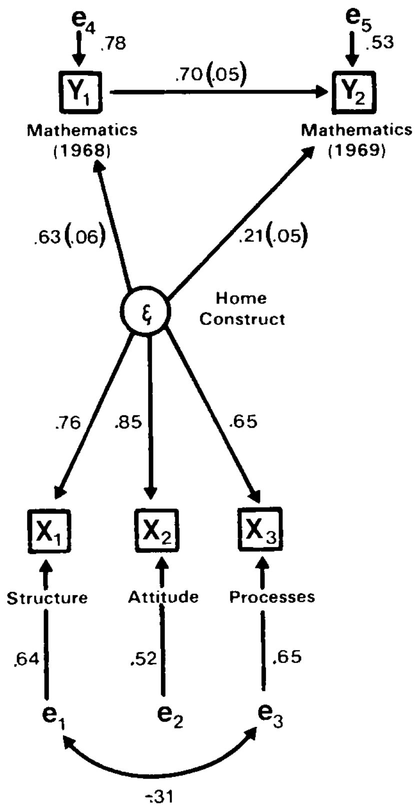
Sörbom and Jöreskog (1981) considered an alternative model that accounted for the presence of measurement error in the home variable. In the structural equation model of Figure 7, the three home variables (X_1 = structure, X_2 = attitude, X_3 = processes) are portrayed as fallible indicators of the "home" construct (ξ). In a factor analytic sense the indicators are said to load on the home factor. The "*e*" terms represent the measurement error associated with the home indicators.

As shown in the diagram, the home variable loadings on the construct are rather high. Thus, the variables provide a reasonable evaluation of the home construct. In contrast to the path analysis result, the effect of home on mathematics is rather significant at both time periods in the LISREL model, when measurement error is eliminated. The illustration demonstrates an instance of when the correction for bias due to measurement error enhanced the interpretation of the results.

In the same model it would probably be useful to allow for measurement error in the endogenous variables mathematics 1968 and mathematics 1969. This is especially true if the achievement measures have low reliabilities. As with the home construct, a reasonable set of achievement measures can be selected to represent the mathematics construct. Often, however, multiple measures are neither available nor practical to collect.

In such a case, measurement error can be accounted for if the reliabilities of the dependent var-

Figure 7
Sörbom-Jöreskog LISREL Model for Home Environment and Mathematics Achievement, with Home as a Latent Variable (Adapted from Sörbom and Jöreskog, 1980, Slightly Modified)



iables are known. A single-indicator structural equation approach can be implemented with the reliabilities used to disattenuate the correlations. The new correlations would then be entered to re-estimate the model. However, the use of multiple indicators is always more effective (Kenny, 1979).

Discussion

Path analysis employs ordinary least-squares techniques to estimate the causal model parameters. If the specified effects are misspecifications, OLS can result in biased estimators. Depending on the seriousness of the misspecifications, erroneous statistical decisions to reject or accept the null hypothesis can result.

In the present discussion, sources of specification error in path analysis models were examined. The empirical illustrations demonstrated the usefulness in formulating and estimating alternative models. A comparison among the model estimates provides a reasonable basis to determine the effects of misspecification in the original theoretical structure.

As shown in the examples, there may be instances where differences between the compared solutions are small. In such instances the alternative structures will, most likely, fail to offer a significant improvement of fit of the original model to the data. On the other hand, extreme differences among the compared solutions suggest a more serious misspecification, given the particular context under investigation, and the alternative is likely to be supported. Thus a given specification error may be serious in one instance but trivial in another context. In any event, the investigator should proceed to consider the tenability of other rival hypotheses about the model's structure. The alternative structures can then be statistically tested for improvement in the fit.

For example, in path analysis models, the test of fit generally employs a comparison between the reproduced correlation matrix (i.e., matrix of zero-ordered correlations resulting from application of the path theorem) and the original matrix. In the event that discrepancies less than .05 occur between the corresponding correlations of the matrices, the model is said to represent a good fit of the given data (Pedhazur, 1982). Often the overall chi-square measure and its associated degrees of freedom and probability level is used to check the model. However, a more detailed assessment of fit is needed to identify which part of the model is misspecified.

Relaxing certain assumptions in path analysis models can result in more complex structural equation models of the LISREL type. It is often reasonable to formulate a hierarchy of alternative theoretical models for comparison. A chi-square difference test provides a useful index for evaluating improvement of model fit (see Bentler & Bonett, 1980.)

Structural equation modeling offers a powerful methodology for investigating relationships among constructs frequently encountered in social science and psychological research. Most importantly it offers a sophisticated technology for examining a wide spectrum of structural hypotheses derived from theory. If used correctly, the results can contribute considerably to efforts to develop and improve social science theories.

References

Bentler, P. Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 1980, *31*, 419–456.

Bentler, P., & Bonett, D. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 1980, *88*, 588–606.

Bielby, W. T., & Hauser, R. M. Structural equation models. *Annual Review of Sociology*, 1977, *3*, 137–161.

Billings, R. S., & Wroten, S. P. Use of path analysis in industrial/organizational psychology: Criticisms and suggestions. *Journal of Applied Psychology*, 1978, *63*, 677–688.

Browne, M. W. Generalized least squares estimates in the analysis of covariance structures. *South African Statistical Journal*, 1974, *8*, 1–24.

Burns, R., & Gallini, J. *A path-analytic study of aptitude-learning relations overtime*. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April 1981.

Cooley, W. W., & Lohnes, P. *Evaluation research in education*. New York: Irvington Publishers, 1976.

Costner, H. L. Theory, deduction, and rules of correspondence. *American Journal of Sociology*, 1969, *75*, 245–263.

Cronbach, L., & Snow, R. E. *Aptitudes and instructional methods*. New York: Irvington Publishers, 1977.

Draper, N. R., & Smith, H. *Applied regression analysis*. New York: Wiley, 1966.

Duncan, O. D. Path analysis: Sociological examples. *American Journal of Sociology*, 1966, *72*, 1–16.

Duncan, O. D. Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin*, 1969, *72*, 177–182.

Duncan, O. L. *Introduction to structural equation models*. New York: Academic Press, 1975.

Gallini, J. Evidence of an adaptive level grading practice through a causal approach. *The Journal of Experimental Education*, 1982, *50*, 188–193.

Goldberger, A. S. Maximum likelihood estimation of regressions containing unobservable independent variables. *International Economic Review*, 1972, *13*, 1–15.

Goodman, L. A. Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology*, 1973, *78*, 1135–1191.

Gough, H. *California psychological inventory*. Palo Alto: Consulting Psychologists Press, 1975.

Heise, D. R. *Causal analysis*. New York: Wiley, 1975.

Jöreskog, K. G. A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar Press, 1973.

Jöreskog, K. G. Structural analysis of covariance and correlation matrices. *Psychometrika*, 1978, *43*, 443–447.

Jöreskog, K. G. Basic issues in the application of LISREL. *DATA*, 1981, *1*, 1–9.

Jöreskog, K. G., & Sörbom, D. *LISREL IV—A general computer program for estimating a linear structural equation system*. Chicago IL: International Educational Services, 1978.

Jöreskog, K. G., & Sörbom, D. *Advances in factor analysis and structural equation models*. Cambridge MA: Abt Associates, Inc., 1979.

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

Keeves, J. P. *Educational environment and student achievement*. Melbourne: Australian Council for Educational Research, 1972.

Kenny, D. *Correlation and causality*. New York: Wiley, 1979.

Kerlinger, F. N., & Pedgazur, E. J. *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston, 1973.

Land, K. C. Principles of path analysis. In E. F. Borgatta (Ed.), *Sociological methodology, 1969:* San Francisco: Jossey-Bass, 1969.

Lee, S., & Bentler, P. Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Journal*, 1980, *14*, 121–136.

Lomax, R. G. A guide to LISREL-type structural equation modelling. *Behavior Research Methods and Instrumentation*, 1982, *14*, 1–8.

Maruyama, G., & McGarvey, B. Evaluating causal models: An application of maximum-likelihood. *Psychological Bulletin*, 1980, *87*, 502–512.

Nygreen, G. T. Interactive path analysis. *The American Sociologist*, 1971, *6*, 37–42.

Pedhazur, E. J. *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston, 1982.

Pindyck, R., & Rubinfeld, D. L. *Econometric model and economic forecasts*. New York: McGraw-Hill, 1981.

Raven, J. Standard progressive matrices. London: H. K. Lewis, 1958.

Snow, R. Learning and individual differences. *Review of Research in Education*, 1976, *4*, 50–105.

Sörbom, D., & Jöreskog, K. G. *The use of structural equation models in evaluation research.* Paper presented at the Conference on Experimental Research in the Social Sciences, Gainesville FL 1981. (preliminary version)

Theil, H. *Principles of econometrics*. New York: Wiley, 1971.

Wolfle, L. M. Strategies of path analysis. *American Educational Research Journal*, 1980, *17*, 183–209.

Werts, C. E., & Linn, R. L. Path analysis: Psychological examples. *Psychological Bulletin*, 1970, *74*, 193–212.

Wright, S. The method of path coefficients. *Annals of Mathematical Statistics*, 1934, *5*, 161–215.

Wright, S. Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics*, 1960, *19*, 189–202.

Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregated bias. *American Statistical Association Journal*, 1962, *57*, 348–368.

Zellner, A., & Theil, H. Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica*, 1962, *30*, 54–78.

Acknowledgments

The author is grateful to an anonymous reviewer whose comments were most useful in preparing the final manuscript. The reported research was in part supported by a Research and Productive Scholarship grant from the Office of Sponsored Research, University of South Carolina.

Author's Address

Send requests for reprints or further information to Joan Gallini, University of South Carolina, Department of Educational Research and Psychology, College of Education, Columbia SC 29208.