DOCUMENT RESUME

ED 384 618 TM 023 260

AUTHOR	Schumacker, Randall E.; Rigdon, Edward
TITLE	Testing Interaction Effects in Structural Equation Models.
PUB DATE	22 Apr 95
NOTE	42p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Computer Software; Computer Software Development; *Interaction; *Path Analysis; Research Problems; *Structural Equation Models
IDENTIFIERS	Latent Variables; LISREL Computer Program

ABSTRACT

Research questions that test the interaction between either categorical variables or multiplicative expressions are commonly found in multiple regression and analysis of variance applications. In contract, research questions posing an interaction between observed variables in a path analytic model are not commonly found in the literature. This is due in part to some of the issues and problems that arise in relational path models. Path models with latent variables also pose problems for the researcher that make the inclusion of interaction tests difficult. This paper further explores the categorical, group, or multi-sample approach and the continuous or multiplicative approach to testing interaction effects in structural equation models that use latent variables. Some software programs have provided a way to test simple interaction effects among latent variables. Examples are included. The developers of structural equation modeling software programs are encouraged to develop procedures for testing interaction effects. Until then, the testing of interaction effects in structural equation models will not be easy to perform. Of the two approaches discussed, multi-sample and multiplicative, the multi-sample approach is generally an easier technique to use for testing the equality of coefficients. An appendix contains a multi-sample interaction example and a LISREL 8 program example. An additional attachment contains other software examples and the outputs of some programs. One table and one figure illustrate the discussion. (Contains 29 references.) (Author/SLD)

TESTING INTERACTION EFFECTS IN STRUCTURAL EQUATION MODELS

Randall E. Schumacker

University of North Texas

Edward Rigdon

Georgia State University

Paper presented at the American Educational Research Association annual meeting

San Francisco, California

April 22, 1995

ABSTRACT

Research questions that test the interaction between either categorical variables or multiplicative expressions are commonly found in multiple regression and analysis of variance applications. In contrast, research questions posing an interaction between observed variables in a path analytic model are not commonly found in the literature. This is due in part to some of the issues and problems that arise in relational path models. Path models with latent variables also pose problems for the researcher that make the inclusion of interaction tests difficult. This paper further explores the categorical, group or **multi-sample** approach and the continuous or **multiplicative** approach to testing interaction effects in structural equation models that use latent variables. Some software programs have provided a way in which to test simple interaction effects among latent variables, and examples are included in the paper. The developers of structural equation modeling software programs are encouraged to further develop procedures for testing interaction effects. Until then, the testing of interaction effects in structural equation models will not be easy to perform. Of the two approaches, multi-sample and multiplicative, the multi-sample approach is generally an easier technique to use for testing the equality of coefficients.

ACKNOWLEDGMENTS

The authors would like to acknowledge the examples provided by Dr. Werner Wothke in Amos, Dr. Michael Neale in Mx , and Dr. Phillip Wood in SAS/Proc Calis using LineEQS commands . Their examples are included in the Appendix of the paper. The program setups and data can be obtained for each as follows:

Mx- via anonymous ftp to opal.vcu.edu in file pub/randy/mod.zip

Amos - via request to smallwaters@acm.org

SAS/Proc Calis via request to wood@psysparc.psyc.missouri.edu.

TESTING INTERACTION EFFECTS IN STRUCTURAL EQUATION MODELS

Research questions that test the interaction between either categorical variables or multiplicative expressions are commonly found in multiple regression and analysis of variance applications. In fact, the rationale for a 2 x 2 fixed-effects factorial study is to first test whether the interaction effect between variables is significant. In the presence of a significant interaction effect, the main effects are not interpreted. The basic reasoning behind testing for interaction is that two variables when interacting produce an effect different from their main effects. For example, hydrogen and oxygen as separate elements (main effects) have specific gaseous states, however, when they interact water is produced. Their interaction produces a different effect than their individual main effects. A rather simplistic example, but it serves to illustrate the basic nature of testing for interaction effects. Since researchers encounter both categorical variables and continuous variables, different approaches have emerged for testing interaction among the variable types. In this paper, the term **multi-sample** will be used when referring to categorical variables and the term **multiplicative** will be used when referring to continuous variables in testing interaction effects.

Types of Interactions

Interaction effects may take many forms. The interaction between two variables may enhance the effect which each variable has separately, as when alcohol and barbiturates are ingested together. In contrast, the interaction may dampen the individual effects of the two variables, as when two noises combine to create a zone of apparent quiet. Smith and Sasaki (1979) describe a "consistency" interaction effect: when the two variables both deviate from their means in the same direction, their joint effect may be enhanced, but if the deviations are in opposite directions, the interaction effect will be opposite, as well. Also, two variables can interact forming a third variable which is entirely different from each variable's individual effect.

Multiple Regression and Analysis of Variance

The testing of interaction effects in multiple regression and analysis of variance have become quite common, and covered in many earlier textbooks (Draper & Smith, 1966; McNeil, Kelly, & McNeil, 1975; Cohen & Cohen, 1975). The reader is probably aware of the various terminology used, namely, polynomial, curvilinear, or non-linear effects used to express a model with an interaction term. A model using continuous variables would take the form:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

where X_3 represents a multiplicative interaction, $X_1{*}X_2$, or could include interaction effects expressed in a model as a polynomial term of the form:

$$Y = \beta_1 X_1 + \beta_2 X_1^2$$

where X^2_1 represents the square of the scores on X_1. More recent treatment of the topic by Aiken and West (1993) further illustrates examples using continuous predictors, categorical predictors, and interactions between categorical and continuous variables.

In general, an interaction effect exists whenever the character of the relationship between variables A and B is affected by the level of variable C (Keppel & Zedeck, 1989). We can demonstrate this by examining a multiple regression equation without interactions:

$$A = \beta_1 * B + \beta_2 * C \qquad (1)$$

The first partial derivatives of this equation with respect to B and C are, respectively, β_1 and β_2.

These derivatives show, for example, that the only effect of a change in B on variable A is the immediate effect through β_1. By contrast, consider a regression equation with an interaction term:

$$A = \beta_1 * B + \beta_2 * C + \beta_3 * (B * C) \qquad (2)$$

Now the first partial derivatives are $\beta_1 + \beta_3 * C$ and $\beta_2 + \beta_3 * B$, respectively. In other words, changes in C, for example, affect A not only directly but also indirectly by changing the character of the relationship between A and B. Accordingly, the estimated direct effect of B on A, β_1, can be thought of as an "average" value computed over the range of values for variable C which are represented in the data set (Keppel & Zedeck, 1989).

Path Analysis Models

Path analysis provides an approach to testing models that indicate direct and indirect effects among specified variable relationships (Pedhazur, 1982; Pedhazur & Schmelkin, 1991). Certain assumptions are made when conducting path analyses beyond those inherent in conducting multiple regression analyses, and it is recommended that these be checked prior to using the technique (Newman, 1991). Other strengths and problems in conducting path analyses have also been elaborated by Pohlmann (1991). Research questions posing an interaction between observed variables in a path analytic model, however, were not commonly found in the research literature (Newman, Marchant, & Ridenour, 1993). Of the studies reporting a test of interaction between variables, the most common approach to test interaction effects was to analyze separate models and compare model weights. In a few studies, a first-order interaction term was included. The authors concluded that, although interaction was implied or explicitly stated in the theory, the models in many of the studies failed to consider interaction effects. Pohlmann (1993) has further indicated that nonlinear relationships and latent variable assumptions can lead to serious specification errors in structural models. Path models with latent variables obviously pose problems for the researcher that make the inclusion of interaction tests difficult.

INTERACTION IN LATENT VARIABLE MODELS

Interaction effects in multiple regression and path analysis with observed variables has posed problems for researchers. These problems with corresponding assumptions are exacerbated in SEM. First, there is the **specification** problem. Linear models ease the task of determining which relationships to investigate and simplify distributional assumptions. Discarding the linearity restriction magnifies the critical role theory plays in focusing the structural equation modeling research effort. Researchers modeling interaction effects must also be sure to collect data which includes a **range of values where interaction effects occur**. Nonlinear functions may appear linear within certain ranges, so it seems prudent to suggest that observed variable values be plotted and investigated prior to model inclusion. In addition, the predictors might be normally distributed, but their **joint-distribution not normally distributed** in a multiplicative interaction term. The test of interaction effects are further complicated by **the degrees of freedom introduced, subsequent interpretation of various fit indices** and **estimated standard errors**, which are only approximate under certain conditions. In the multiplicative model when multiplying two observed variables, the product will contain the "true scores"and error with the distribution of the **products of the error variables not normally distributed** even if the error variables themselves are These concerns, as well as, several others discussed in the paper complicate testing interaction effects in structural equation models.

There are two general approaches for dealing with interaction effects in structural equation models. Interactions involving discrete or discretized (categorical) variables can be evaluated using multi-sample analysis, or researchers can model interactions between two continuous variables by creating measures of the interaction "construct," using techniques similar to those pioneered by Kenny and Judd (1984). Each of these techniques has its own strengths and weaknesses, and will be discussed separately. The multi-sample approach is more readily adaptable to testing the equality of factor structures, equal regression coefficients, and means of latent variables (Jöreskog & Sörbom, 1993a).

Multi-Sample Approach

In the multi-sample approach, the different samples are defined by the different levels of one or both of the interacting variables. If interaction effects are present, then certain parameters should have different values in the different samples. For example, in the context of an experiment, Mackenzie and Spreng (1992) tested an interaction model of advertising effectiveness. Their model proposed, in part, that a person's attitude toward an advertisement (A_a) interacted with the person's motivation to process the ad to affect the person's attitude toward the advertised brand (A_b). The structural model may be represented as:

$$\eta_1 = \alpha_1 + \gamma_{11} * \xi_1 + \zeta_1 \qquad (3)$$

where η_1 is A_a, α_1 is the intercept, ξ_1 is A_b, and ζ_1 is an error term. Mackenzie and Spreng hypothesized that subjects with a higher motivation to process the ad would have a higher mean level of A_b after viewing the ad. This would be reflected in a higher intercept term for the subjects in the high motivation condition. They proposed that A_a should have a positive "main" effect (as represented by γ_{11}) on A_b. However, they further proposed that the interaction between A_a and motivation would make the A_a--A_b relationship weaker in the high motivation condition (see Figure 1).

Insert Figure 1 about here

Under the multi-sample approach, researchers can investigate interaction effects using χ^2 difference tests. In this example, the two samples are composed of subjects who were either exposed or not exposed to a manipulation designed to increase motivation. To test for the interaction between motivation and A_a, first estimate a model where γ_{11} is restricted to be equal across the two groups, and then estimate a model where the parameter is allowed to differ in the two samples. The two models are nested, so that, given assumptions, the test statistic is a χ^2

difference test with one degree of freedom. In the example drawn from Mackenzie and Spreng (1992), the χ^2 statistics for the two models were 44.55 (25 df) and 36.38 (24 df). The difference, 8.17 (1 df) is significant ($p < .005$), so the interaction hypothesis is supported. The maximum likelihood estimate for γ_{11} was 0.881 ($t = 13.99$) for the low motivation subjects but only 0.644 ($t = 9.58$) for the high motivation subjects. The Appendix contains the LISREL 8 command file for this problem.

The multi-sample approach is the natural choice when the interaction relationship involves one or more categorical variables, and this approach has additional strengths as well. It can be used to represent a very wide variety of interaction effects without requiring substantial new methodological developments. Even the extension to higher order interaction effects is fairly obvious. This approach can be used regardless of whether the interaction intensifies or mutes the effects of the individual variables. Because the interaction effect is represented in the difference between samples, the researcher may be able to preserve linear relations between variables within each sample, thus avoiding potentially significant complications in fitting the model. Finally, nearly every commercially available SEM software package allows for multi-sample analysis with restrictions across samples, so researchers can probably use the package with which they are most familiar.

The multi-sample approach also has a number of weaknesses. First, this method forces the researcher to divide the total available sample size by the number of groups. In a methodology which relies so heavily on asymptotic properties, and where those properties may only be achieved at relatively high sample sizes, this is a serious limitation. Furthermore, this reduction in sample size may confound results from χ^2 difference tests. MacCallum, Roznowski and Necowitz (1992) noted a substantial degree of instability of fit indices in even moderately sized samples. Thus, it is possible that the multi-sample approach may yield subsamples which are so small that instability in the χ^2 statistic will mislead the researcher into believing that an interaction effect exists, whether it does or not. On the other hand, the researcher may be able to minimize this problem by estimating only those distinct parameters that are necessary. In particular, researchers usually have no reason to expect that measurement parameters, such as loadings, will be different in the various subsamples. Specifying equality for such parameters may partially lessen the impact of sample division on results.

Additional problems arise if researchers attempt to apply the multi-sample approach to interactions involving two continuous variables. In such cases, the researcher must reduce at least one of the continuous variables to a categorical variable, for purposes of defining the groups. The categorized variable will contain less information than the original continuous variable, and loss of information is undesirable (Russell & Bobko, 1992). Furthermore, in the process of categorization, a number of observations will probably be misclassified due to measurement error. SEM researchers account for measurement error by using multiple measures and latent variables, but latent variables are typically modeled as continuous. Perhaps researchers could compute factor scores and perform the transformation on that basis. This approach however has not been reported in the literature. Furthermore, choosing the point(s) at which to divide the sample into the various groups is(are) arbitrary. If researchers use arbitrary values (such as the median), mere random sampling error will ensure that many cases are misclassified, violating the basic assumption that the cases in a particular subsample all come from the same population.

The Continuous-Variable Approach

When an interaction involves two continuous variables, researchers may wish to adapt procedures similar to those described by Kenny and Judd (1984). In the continuous-variable approach, an interaction construct, for example ξ_3, is represented explicitly, along with ξ_1 and ξ_2, the "main effect" constructs, in a structural model of the form:

$$\eta_1 = \gamma_1 * \xi_1 + \gamma_2 * \xi_2 + \gamma_3 * \xi_3 + \zeta_1 \tag{4}$$

where η_1 is the dependent construct and ζ_1 is an error term. Notice that the relationship between η_1 and ξ_3 is itself linear. Thus the researcher must create ξ_3 so as to represent the kind of nonlinear relationship believed to exist between η_1 and ξ_1 and ξ_2. For example, to check for a multiplicative interaction between ξ_1 and ξ_2, ξ_3 must be made equal to $\xi_1 * \xi_2$. Measures or indicators of ξ_3 are then created as functions of the measures of ξ_1 and ξ_2, in a similar way.

The structure of the interaction model emerges as a logical extension of the measurement model for ξ_1 and ξ_2. It should be noted that these developments are due to the pivotal work by Kenny and Judd (1984). The basic factor analysis model can be stated as:

$$\mathbf{x} = \mathbf{\Lambda} * \boldsymbol{\xi} + \boldsymbol{\delta} \tag{5}$$

where $\mathbf{x}$ is a vector of manifest variables, $\mathbf{\Lambda}$ is a matrix of loadings, and $\boldsymbol{\delta}$ is a vector of measurement error terms. In thinking about the continuous-variable approach, it will help to keep a number of points in mind. First, remember that both ξ and $\boldsymbol{\delta}$ are latent variables affecting $\mathbf{x}$. While researchers may habitually draw sharp mental distinctions between the two, there is little difference between them, beyond the basic fact that one set of latent variables are common to several manifest variables while the other set load uniquely on individual manifest variables.

Second, remember the assumptions underlying the basic factor model. The common factors, $\boldsymbol{\xi}$, are assumed to be uncorrelated with the unique factors, $\boldsymbol{\delta}$. Both the common and unique factors are assumed to be normally distributed and have zero means. We can represent the covariance matrices of the common and unique factors as $\mathbf{\Phi}$ and $\mathbf{\Theta}$, respectively. In modeling a multiplicative interaction, Kenny and Judd (1984) cited a third useful result from the theory of normally distributed variables. If A and B are normally distributed, then:

$$\sigma^2(A * B) = \sigma^2(A) * \sigma^2(B) + (\sigma(A, B))^2 \tag{6}$$

That is, the variance of the product is equal to the product of the variances plus the square of the covariance. Obviously, if A and B are uncorrelated, then the last term disappears.

From this point, Kenny and Judd (1984) used simple algebraic substitution to develop their model of multiplicative interaction effects. Suppose there are two measures which follow factor models:

$$x_1 = \lambda_1 * \xi_1 + \delta_1 \tag{7}$$

$$x_2 = \lambda_2 * \xi_2 + \delta_2 \tag{8}$$

Then the product $\mathbf{x}_3 = \mathbf{x}_1 * \mathbf{x}_2$ should be reflected in the following model as:

$$x_3 = \lambda_1 * \lambda_2 * \xi_1 * \xi_2 + \lambda_1 * \xi_1 * \delta_2 + \lambda_2 * \xi_2 * \delta_1 + \delta_1 * \delta_2 \qquad (9)$$

Or

$$x_3 = \lambda_3 * \xi_3 + \lambda_1 * \xi_4 + \lambda_2 * \xi_5 + \delta_3 \qquad (10)$$

where

$$\xi_3 = \xi_1 * \xi_2 \qquad (11)$$

$$\xi_4 = \xi_1 * \delta_2 \qquad (12)$$

$$\xi_5 = \xi_2 * \delta_1 \qquad (13)$$

$$\delta_3 = \delta_1 * \delta_2 \qquad (14)$$

$$\lambda_3 = \lambda_1 * \lambda_2 \qquad (15)$$

All of these new latent variables are mutually uncorrelated and uncorrelated with all other latent variables in the model.

In order to incorporate this interaction effect into a model, Kenny and Judd (1984) needed to specify $\mathbf{x}_3$ as a function of latent variables whose variances and covariances reflected these relationships. This involved specifying some model parameters as nonlinear functions of other parameters. Kenny and Judd (1984) used a version of the COSAN program (Fraser, 1980) to accomplish this. Hayduk (1987), applying Rindskopf's (1984) work with "phantom" constructs, showed how the Kenny-Judd model could be specified artificially using the simple equality constraints found in programs like Jöreskog and Sörbom's LISREL software. However, the difficulty of using Hayduk's method was sufficient to dissuade most researchers from attempting to specify such an interaction effect in a model. Today, both SAS/Proc CALIS (1990) (cited by Waller, 1993; program by Wood in Appendix) and Jöreskog and Sörbom's (1993b) LISREL 8, as well as, Amos and Mx software programs (see Appendix), include the ability to specify nonlinear constraints directly.

LISREL8 EXAMPLE

With LISREL 8, researchers can specify the Kenny and Judd (1984) interaction model using VAlue and EQuality commands and the new COnstraint commands. For example, the ı .ny-Judd model implies that:

$$\sigma^2(\xi_3) = \sigma^2(\xi_1) * \sigma^2(\xi_2) + (\sigma(\xi_1, \xi_2))^2 \quad (16)$$

Researchers can specify this relationship using the COnstraint line:

CO PH(3,3) = PH(1,1) * PH(2,2) + PH(2,1)**2

Similarly, the model implies that:

$$\sigma^2(\xi_4) = \sigma^2(\xi_1) * \sigma^2(\delta_2) \quad (17)$$

and this relationship can be specified via:

CO PH(4,4) = PH(1,1) * TD(2,2)

One peculiarity of LISREL 8 is that all terms on the right side of a CO statement must be free parameters--they cannot be fixed values and they cannot themselves be subject to constraints (Jöreskog & Sörbom, 1993b). If a parameter is fixed, then the researcher should replace the parameter in the COnstraint line with the fixed value. If the parameter is constrained, then substitution may eliminate the problem. Furthermore, LISREL 8's routine for computing starting values will not work for the continuous-variable interaction model. The starting values chosen by the researcher can have a tremendous impact on the ability of the software to converge on a solution, even when the model is correct in the population.

Kenny and Judd (1984) demonstrated their approach by fitting a covariance matrix computed from 500 Monte Carlo observations drawn from a population defined by an interaction model with known parameter values. A complete LISREL 8 program specification for the Kenny and Judd model is in the Appendix. Table 1 reports the true parameter values and the estimates produced by Kenny and Judd (1984), Hayduk (1987), and the LISREL 8 program in the Appendix. All three sets of results are very similar to the true parameter estimates, and to each other.

Insert Table 1 about here

Researchers should consider using the continuous-variable approach particularly when wanting to avoid the weaknesses of the multi-sample approach which included: splitting the sample and possibly misclassifying observations; categorizing a variable and losing information; and increased sample size requirements. Kenny and Judd (1984) noted a possible benefit of using the continuous-variable approach, namely, parsimony. All but one of the additional parameters involved in the interaction model are exact functions of the "main effects" parameters. The only new parameter is the γ parameter linking the interaction construct to the dependent construct.

However, the continuous-variable approach has its own weaknesses, and specifying a continuous-variable interaction model is still a tedious, complex undertaking, with a high risk of making programming errors. Second, if a model includes several measures of a construct, or higher-order interaction terms, the technique becomes even more difficult to program. If constructs A and B have **a** and **b** measures, respectively, then the interaction construct AB will have **a * b** measures. If each construct has four measures, as is necessary to apply Anderson and Gerbing's (1988) internal consistency criterion, then the interaction construct will have 16 measures. With the four measures of the two "main effect" constructs, and at least one measure of the dependent construct, the model will have at least 25 manifest variables before considering any other constructs.

Under the continuous-variable approach, the researcher must specify the functional form for the interaction, construct measures of the interaction construct, and then model the relationships between the interaction measures and the latent variables. The Kenny-Judd model involved a specific type of multiplicative interaction, but this is hardly the only form that an interaction can take. For other types of interaction there is little prior work available to guide researchers. Researchers may also face a profound multicollinearity problem. It is very likely that the interaction measures will be highly correlated with the measures used to construct them. This multicollinearity can be devastating to factor analytic measurement models, causing measures to be more highly correlated with measures of other constructs than they are with measures of the same construct. For multiplicative interactions betw en normally distributed variables, Smith and Sasaki (1979) demonstrated that researchers can eliminate this multicollinearity by centering the original variables--expressing them as deviations from their means--before computing the product variables. However, centering the variables alters .ie form of the interactive relationship. Notice that the product variable A * B will have the same value when A and B are both well above their means and when they are both well below their means. Thus, Smith and Sasaki (1979) noted that centering the variables turns the original multiplicative interaction into a kind of "consistency effect." Researchers who want to model other types of interactions may find no easy answer to the problem of multicollinearity.

The interaction model also presents distributional problems which are much more serious than those associated with SEM techniques in general. First, the formula for the variance of a product variable (Equation 6) assumes that the two original variables are normally distributed. If the original variables are severely nonnormal, the variance of the product variable can be very different from the value implied by Equation 5, and the interaction model will perform poorly. Of course, suitable data transformations may result in approximate normal distributions for the original variables. There is, however, a second distributional problem. Many, if not most, functions of two normally distributed variables will not themselves be normally distributed. This problem applies to both the measures and the latent variables. This nonnormality violates distributional assumptions associated with estimation methods such as maximum likelihood. Furthermore, estimation methods that do not make distributional assumptions may not work for interaction models. Jöreskog and Yang (1995) indicated that the asymptotic weight matrix associated with the covariance matrix for an interaction model may be not positive definite, due to dependencies between moments of different variables which are implied by the interaction model. Distribution-free estimation methods that use the inverse of the asymptotic covariance matrix will therefore fail. Researchers may minimize this problem by estimating the model using a distribution-sensitive method but evaluating the results with a bootstrapping technique. The rescaled bootstrapping method outlined by Bollen and Stine (1993) may be particularly appropriate in this case. Note that techniques for computing bootstrap estimates of the weight matrix will not be helpful in this case, because the weight matrix is singular in the population (Yung & Bentler, 1994).

CONCLUSION

Clearly, both the multi-sample and continuous variable methods have their weaknesses. While new developments in SEM software are creating exciting new possibilities, researchers need to remain focused on their research hypotheses and theory. The multi-sample approach appears to be the more generally useful of the two techniques and gives rise to fewer problems, if theory, hypotheses, and variables are appropriately considered. Researchers who hypothesize "consistency effect" interactions are probably advised to apply the Kenny and Judd (1984) approach, but for other researchers, the difficulties of the continuous variable approach may not be worth the effort. Researchers may balk at the sample-splitting that is a part of the multi-sample approach, but in our experience, this problem is only a major concern in cases where researchers are working with data sets that were not collected with an interaction model in mind. Researchers will always face significant problems when they attempt to use data for unintended purposes, especially in SEM. Furthermore, it is not clear, in general, which of the two approaches described here is more robust to this kind of ex post facto modeling. Overall, we encourage further development of the software programs to more readily incorporate tests of interaction effects and address many of the issues and concerns indicated by the research literature.

REFERENCES

Aiken, L.S. & West, S.G. (1993). Multiple regression: Testing and interpreting interactions. SAGE Publications: Newbury Park, CA.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin, 103(3), 411-423.

Bollen, K. A., & Stine, R. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 111-135). Newbury Park, CA: Sage.

Cohen, J. & Cohen, P. (1975). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, Publishers: Hillsdale, NJ.

Draper, N.R. & Smith, H. (1966). Applied Regression Analysis. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc.: New York, NY.

Fraser, C. (1980). COSAN user's guide. The Ontario Institute for Studies in Education: Toronto, CA

Hayduk, L. A. (1987). Structural equation modeling with LISREL: Essentials and advances. Baltimore: The Johns Hopkins University Press.

Jöreskog, K. G., & Sörbom, D. (1989). LISREL 7: A Guide to the program and applications (2nd ed.). Chicago: SPSS.

Jöreskog, K. G., & Sörbom, D. (1993a). LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Scientific Software International: Chicago, IL.

Jöreskog, K. G., & Sörbom, D. (1993b). LISREL 8 user's reference guide. Chicago: Scientific Software.

Jöreskog, K. G., & Yang, F (1995). Non-linear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), Advanced structural equation modeling techniques. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin, 96(#July), 201-210.

Keppel, G., & Zedeck, S. (1989). Data analysis for research designs: Analysis of variance and multiple correlation/regression approaches. New York: W. H. Freeman.

Lilien, G. L. & Kotler, P. (1983). Marketing decision making: A model-building approach. New York: Harper & Row.

Mackenzie, S. B., & Spreng, R. A. (1992). How does motivation moderate the impact of central and peripheral processing on brand attitudes and intentions. Journal of Consumer Research, 18(March), 519-529.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. Psychological Bulletin, 111(May), 490-504.

McNeil, K.A., Kelly, F.J., & McNeil, J.T. (1975). Testing Research Hypotheses Using Multiple Linear Regression. Southern Illinois University Press: Carbondale, IL.

Newman, I. (1991, April). To path analyze or not to path analyze: Is there an alternative approach. Paper presented at American Educational Research Association, Chicago, Illinois.

Newman, I., Marchant, G.J., Ridenour, T. (1993, April). Type VI errors in path analysis: Testing for interactions. Paper presented at American Educational Research Association, Atlanta, Georgia.

Pedhazur, E.J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd Ed.). CBS College Publishing: New York, NY.

Pedhazur, E.J. & Schmelkin, L.P. (1993). Measurement, design, and analysis: An integrated Approach. Lawrence Erlbaum Associates, Publishers: Hillsdale, NJ.

Pohlmann, J. (1991, April). Issues in path analysis: Strengths and problems. Paper presented at American Educational Research Association, Chicago, Illinois.

Pohlmann, J. (1993). Insidious structural errors in latent variable models. Mid-Western Educational Researcher, 6(3), 29-32.

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. Psychometrika, 49(March), 37-47.

Russell, C. J., & Bobko, P. (1992). Moderated regression analysis and likert scales: Too coarse for comfort. Journal of Applied Psychology, 77(June), 336-42.

SAS. (1990). User's guide, version 6, volume 1 (4th ed.). Cary, NC:

Smith, K. W., & Sasaki, M. S. (1979). Decreasing multicollinearity: A method for models with multiplicative functions. Sociological Methods & Research, 8(August), 35-56

Waller, N. G. (1993). Seven confirmatory factor analysis programs: EQS, EzPath, LINCS, LISCOMP, LISREL 7, SIMPLIS, and CALIS. Applied Psychological Measurement, 17(March), 73-100.

Yung, Y.-F., & Bentler, P. M. (1994). Bootstrap-corrected ADF test statistics in covariance structure analysis. British Journal of Mathematical and Statistical Psychology, 47(May), 63-84.

APPENDIX

MULTI-SAMPLE INTERACTION EXAMPLE

(Annotated)

LISREL8 command lines are flush left. Comments are indented.

When estimating multi-sample problems in LISREL, it is important to include a different title card for each group, in order to make the output more readable.

TITLE CARD: Low Motivation Sample

> The DA card includes the NG parameter, which specifies the number of groups or samples. As in all multi-sample problems, researchers should be careful to analyze the covariance matrix rather than the correlation matrix (Jöreskog & Sörbom, 1989).

DA NI=6 NO=200 MA=CM NG=2

> This problem requires both covariances and means. These keywords would be followed by either the moments themselves or by their file locations.

CM

ME

LA; ATT_AD1 ATT_AD2 ATT_AD3 BRANDAT1 BRANDAT2 BRANDAT3

> Mackenzie and Spreng (1992) specified their model in terms of η constructs and y variables only. While it is not required for current purposes, it does simplify the programming.

MO NE=2 NY=6 LY=FU,FI BE=FU,FI PS=SY TE=SY TY=FR AL=FI

LE; ATT_AD BRANDATT

FR LY(2,1) LY(3,1) LY(5,2) LY(6,2)

VA 1 LY(1,1) LY(4,2)

FR BE(2,1)

OU AD=OFF

> As Jöreskog and Sörbom (1989) note, setting up multi-sample problems with LISREL is a less onerous task than one might expect. Specifications for the first group in the set-up become default values for later groups.

TITLE CARD: High Motivation Sample

DA NO=160

CM

ME

LA; ATT_AD1 ATT_AD2 ATT_AD3 BRANDAT1 BRANDAT2 BRANDAT3

> The construct intercepts (the α's) are only defined relatively across the two groups. Thus, the intercepts for the first group are fixed to 0. Then estimates of the intercepts in the second group are evaluated relative to 0.

MO LY=IN BE=PS TE=PS PS=PS TY=IN AL=FR

OU

LISREL 8 PROGRAM FOR THE KENNY-JUDD MODEL

Program lines are flush left.

Comments are indented one-half inch.

Title card: Kenny-Judd (1984) simple example with fewest starting values

DA NI=9 NO=500 MA=CM

> Kenny and Judd labeled their independent constructs X and Z, and labeled their measures similarly. The underlines in these labels indicate multiplication.

LA; x1 x2 z1 z2 x1_z1 x1_z2 x2_z1 x2_z2 y

SE; 9 1 2 3 4 5 6 7 8

CM

> For most parameter matrices, most elements will be fixed, so it makes sense to begin with these matrices set to FIxed.

MO NY=1 NE=1 NX=8 NK=7 PS=SY TE=FI PH=SY,FI TD=SY,FI GA=FI

> This program includes ETA and KSI in the latent variable labels in order to minimize confusion between measures and constructs. Note that, for the KSI latent variables, 1 and 2 are the main effects, X and Z, while 3 is the interaction, X*Z.

LE; ETA_Y

LK; KSI_X KSI_Z KSI_XZ X_TD3 X_TD4 Z_TD1 Z_TD2

> These fixed loadings set the scales of constructs Y, X and Z.

VA 1.0 LY(1,1) LX(1,1) LX(3,2)

> These are the free parameters relating to the main effect constructs and their measures.

FR GA(1,1) GA(1,2) PH(1,1 PH(2,2) PH(2,1)

FR LX(2,1) LX(4,2) TD(1,1) TD(2,2) TD(3,3) TD(4,4)

> The direct effect of the interaction construct on the dependent consruct is the only additional free parameter in the interaction model.

FR GA(1,3)

> Following this are the constrained elements of the interact model. The loadings of the interaction measures on the interaction construct are simple products of the loadings of x1 and x2 on X and z1 and z2 on Z. Since two of those loadings are fixed to 1, they do not appear in the COnstraint statements. Remember that the parameters that actually appear on the right side of a COnstraint line must themselves be entirely FRee.

VA 1.0 LX(5,3)

CO LX(6,3) = LX(4,2)

CO LX(7,3) = LX(2,1)

CO LX(8,3) = LX(2,1) * LX(4,2)

Note that this program puts the measurement error variances into Θ_δ. Alternately, the program could have created more ξ constructs to represent these--but that would needlessly increase the dimension of several parameter matrices, slowing down the software. With 8 exogenous measures, Θ_δ will be 8 x 8 anyway, so why not put the product measurement error variances there? Note that the covariances of these measurement error terms are all zero. That is why those covariances do not appear in these COnstraint lines.

```
CO TD(5,5) = TD(1,1) * TD(3,3)
CO TD(6,6) = TD(1,1) * TD(4,4)
CO TD(7,7) = TD(2,2) * TD(3,3)
CO TD(8,8) = TD(2,2) * TD(4,4)
```

On the other hand, this program specifies that the constructs X and Z are correlated, so we must remember their covariance in specifying the variance of X*Z.

```
CO PH(3,3) = PH(1,1) * PH(2,2) + PH(2,1)**2
```

And these constraints represent the loadings of the interaction measures on the nuisance constructs, the products of constructs and error terms. In looking at these constraints, remember that, in LISREL 8, the loading of a measure on its own measurement error is implicitly fixed to 1. Also remember that LX(1,1) and LX(3,2) are fixed to 1. All of those fixed 1's are 'invisible' elements of these constraints.

```
VA 1.0 LX(5,4)
CO LX(7,4) = LX(2,1)
CO PH(4,4) = PH(1,1) * TD(3,3)
VA 1.0 LX(6,5)
CO LX(8,5) = LX(2,1)
CO PH(5,5) = PH(1,1) * TD(4,4)
VA 1.0 LX(5,6)
CO LX(6,6) = LX(4,2)
CO PH(6,6) = PH(2,2) * TD(1,1)
VA 1.0 LX(7,7)
CO LX(8,7) = LX(4,2)
CO PH(7,7) = PH(2,2) * TD(2,2)
```

The Kenny-Judd model will run in LISREL 8 with only these starting values specified. In some cases, the problem will require starting values for virtually every parameter in the model. Do not ignore the constrained parameters, either. Just because the 'original' parameters have starting values, LISREL 8 will not automatically compute starting values for constrained parameters on that basis. The NS parameter on the output line suppresses automatic starting value calculation--this may eliminate the occasional processing error.

```
ST -.15 GA(1,1)
ST .35 GA(1,2)
ST .70 GA(1,3)
```

LISREL 8's admissibility check will be unhelpful for interaction models, and should be turned off. ML is LISREL's default fit function, but it is specified here for clarity.

```
OU ML AD=OFF
```

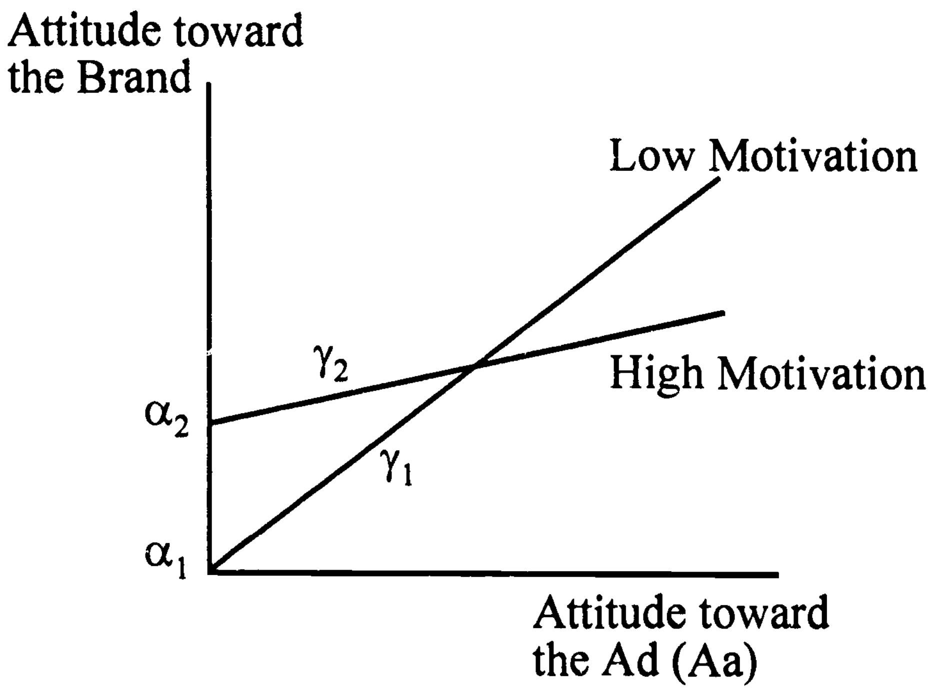
Table 1

Comparison of Parameter Estimates

		Estimates		
Parameter	Population Value	Kenny-Judd (1984)	Hayduk (1987)	LISREL 8
γ_{11}	-.15	-.17	-.17	-.17
γ_{21}	.35	.32	.32	.32
γ_{31}	.70	.70	.71	.71
λ_{21}	.60	65	.64	.65
λ_{42}	.70	.69	.69	.69

Figure Captions

Figure 1. An example of an interaction effect.



γ_1 : Main effect of Aa
$\alpha_2 - \alpha_1$: Main effect of Motivation
$\gamma_2 - \gamma_1$: Interaction effect

MX SOFTWARE EXAMPLES

mod.sum:

Moderator variables: two approaches.

The standard approach allows for main effects of a predictor variable and a moderator variable, along with a multiplicative interaction term predictor*moderator. This may be modelled with either summary statistics such as means and covariances, or with the raw data.

The Mx program, moda.mx, illustrates this procedure with simulated data on 500 subjects. A structural equation model of multiple regression of the outcome (Y) variable on the three X variables is fitted by maximum likelihood to the raw data vectors. The last two lines of the script revise it to drop out the regression on predictor*moderator. The difference in the fit function between these two runs is 120.6 which is distributed as chi-square with 1 d.f. (i.e., 5270.752 - 5150.144 = 120.6).

The Mx program, modb.mx shows an alternative approach to the problem. Rather than pre-compute the moderator*predictor variable and use the multiple regression model, we use special features in Mx to place the observed moderator values directly on a path in the diagram. This time we are still regressing the outcome variable on the predictor and moderator main effects, but a second copy of the moderator variable is used as a path in the structural equation model. During model-fitting, Mx is recomputing the predicted covariance for each subject, so that every subject has a different model based on their moderator variable value. Again we fit the model with and without the moderator effect, and we recapture the same difference in log-likelihood between the two runs with 1 d.f. (i.e., 3698.475 - 3577.868 = 120.6).

This method takes quite a bit longer to run than the first approach, so what is its value? Consideration of the further possibilities of this method gives the answer. We could very easily revise the script to employ some arbitrarily complex function of the moderator variable. Perhaps we expect some exponential model? Or a sinusoidal one? Whatever theoretical expectations we have, they can be modelled directly and uniquely for each subject.

Moda.mx Program

```
moda.mx:
Title Moderated regression with raw data
Data Ninput=4 Nobs=500 Ngroups=1
Rectangular file=moda.raw ! Data file contains 4 variables:
Labels Predictor Moderator Predictor*Moderator Outcome
Matrices
R symm 3 3 Free  !Cov matrix of Pred, Mod, Pred*Mod
L full 1 3 Free  !Regression of Outcome on above 3 variables
D diag 1 1 Free  !Error of outcome
M full 1 4 Free  !Matrix for estimated means of all 4 variables
Means M;         !Predicted means
Cov R | R*L' _
 L*R | L*R*L'+D ;  !Predicted covariance structure
Start .5 all       !Starting values
Start 3 R 1 1 R 2 2 R 3 3
Option multiple-fit !To allow fitting of submodel after the end statement
End
Drop 9  !Fix at zero the 9th parameter, regression of outcome on pred*mod
End
```

Modb.mx Program

```
modb.mx:
Title Moderated regression with raw data
#define nx 2  ! Number of X variables
#define ny 1  ! Number of Y variables
Data Ninput=4 Nobservations=500 Ngroups=1
Labels Predictor ModeratorMainEffect Moderator Outcome
Rectangular file=modb.raw
Definition Moderator /
Matrices
R symm nx nx Free !Cov matrix of Predictor and ModeratorMainEffect (X vars)
J full ny nx Free !Regression of Outcome on X variables
K full ny nx     !To contain individual moderator values for each subject
L full ny nx     !Free parameter for magnitude of moderating effect
D diag ny ny Free !Error of Outcome (Y)
M full 1 3 Free  !For estimated means of X & Y
Means M;         !Matrix expression for means
Cov R | R*(J+L.K)' _
 (J+L.K)*R | (J+L.K)*R*(J+L.K)'+D ; !Matrix expression for covariance
Icodes 1 2 4 99  !To indicate that variables 1, 2 & 4 are being analyzed
Specify K -1 0   !Individual moderator values will go in K
Specify L 100 0  !Free parameter for size of moderator effect
Start .5 all
Start .0 L 1 1
Start 1 R 1 1 R 2 2  !Starting values
Option Multiple-fit  !To allow fitting of submodel
End
drop -1 100      !Refit model with moderator effect fixed at zero
End
moda.raw: <insert raw data file here from SAS program via ftp address>
modb.raw: <insert raw data file here from SAS program via ftp address>
```

OUTPUT OF MODA.MX
(abridged)

```
** Mx startup successful **
** MX-Windows version 1.30**

Summary of VL file data for group        1

   Code       1.0000    2.0000    3.0000    4.0000
 Number     500.0000  500.0000  500.0000  500.0000
   Mean       0.0314   -0.0189    0.5145    0.1258
Variance      1.0162    1.0074    1.3608    1.0690

MX PARAMETER ESTIMATES

GROUP NUMBER:       1
 TITLE MODERATED REGRESSION WITH RAW DATA

 MATRIX D
This is a DIAGONAL matrix of order    1 by    1
       1
1   0.3392

 MATRIX L
This is a FULL matrix of order    1 by    3
       1       2       3
1   0.4972  0.4281  0.2609

 MATRIX M
This is a FULL matrix of order    1 by    4
       1       2       3       4
1   0.0314 -0.0189  0.5145  0.1258

 MATRIX R
This is a SYMMETRIC matrix of order    3 by    3
          1          2          3
1     1.0162
2     0.5151     1.0074
3    -0.0420    -0.0318     1.3608

 Your model has   14 estimated parameters and   2000 Observed statistics
```

-2 times log-likelihood of data = 5150.144

Multiple fit option in effect

The following MX script lines have been read:

DROP 9
END

Summary of VL file data for group 1

Code	1.0000	2.0000	3.0000	4.0000
Number	500.0000	500.0000	500.0000	500.0000
Mean	0.0314	-0.0189	0.5145	0.1258
Variance	1.0162	1.0074	1.3608	1.0690

MX PARAMETER ESTIMATES

GROUP NUMBER: 1
TITLE MODERATED REGRESSION WITH RAW DATA

MATRIX D

This is a DIAGONAL matrix of order 1 by 1

	1
1	0.4317

MATRIX L
This is a FULL matrix of order 1 by 3

	1	2	3
1	0.4883	0.4245	0.0000

MATRIX M
This is a FULL matrix of order 1 by 4

	1	2	3	4
1	0.0314	-0.0189	0.5145	0.1258

MATRIX R
This is a SYMMETRIC matrix of order 3 by 3

	1	2	3
1	1.0162		
2	0.5151	1.0074	
3	-0.0420	-0.0318	1.3608

Your model has 13 estimated parameters and 2000 Observed statistics

-2 times log-likelihood of data = 5270.752

OUTPUT OF MODB.MX

** Mx startup successful **
MX-Windows version 1.30

Summary of VL file data for group 1

Code	-1.0000	1.0000	2.0000	4.0000
Number	500.0000	500.0000	500.0000	500.0000
Mean	-0.0189	0.0314	-0.0189	0.1258
Variance	1.0074	1.0162	1.0074	1.0690

MX PARAMETER ESTIMATES

GROUP NUMBER: 1
TITLE MODERATED REGRESSION WITH RAW DATA

MATRIX D
This is a DIAGONAL matrix of order 1 by 1

	1
1	0.3392

MATRIX J
This is a FULL matrix of order 1 by 2

	1	2
1	0.4972	0.4363

MATRIX K
This is a FULL matrix of order 1 by 2

	1	2
1	0.7378	0.0000

MATRIX L
This is a FULL matrix of order 1 by 2

	1	2
1	0.2609	0.0000

MATRIX M
This is a FULL matrix of order 1 by 3

	1	2	3
1	0.0314	-0.0189	-0.0086

MATRIX R
This is a SYMMETRIC matrix of order 2 by 2

	1	2
1	1.0162	
2	0.5151	1.0074

Your model has 10 estimated parameters and 2000 Observed statistics

-2 times log-likelihood of data = 3577.868

Multiple fit option in effect

The following MX script lines have been read:

DROP -1 100
END

Summary of VL file data for group 1

Code	-1.0000	1.0000	2.0000	4.0000
Number	500.0000	500.0000	500.0000	500.0000
Mean	-0.0189	0.0314	-0.0189	0.1258
Variance	1.0074	1.0162	1.0074	1.0690

MX PARAMETER ESTIMATES
GROUP NUMBER: 1
TITLE MODERATED REGRESSION WITH RAW DATA

MATRIX D
This is a DIAGONAL matrix of order 1 by 1

	1
1	0.4317

MATRIX J
This is a FULL matrix of order 1 by 2

	1	2
1	0.4883	0.4245

MATRIX K
This is a FULL matrix of order 1 by 2

	1	2
1	0.7378	0.0000

MATRIX L
This is a FULL matrix of order 1 by 2

	1	2
1	0.0000	0.0000

MATRIX M
This is a FULL matrix of order 1 by 3

	1	2	3
1	0.0314	-0.0189	0.1258

MATRIX R
This is a SYMMETRIC matrix of order 2 by 2

	1	2
1	1.0162	
2	0.5151	1.0074

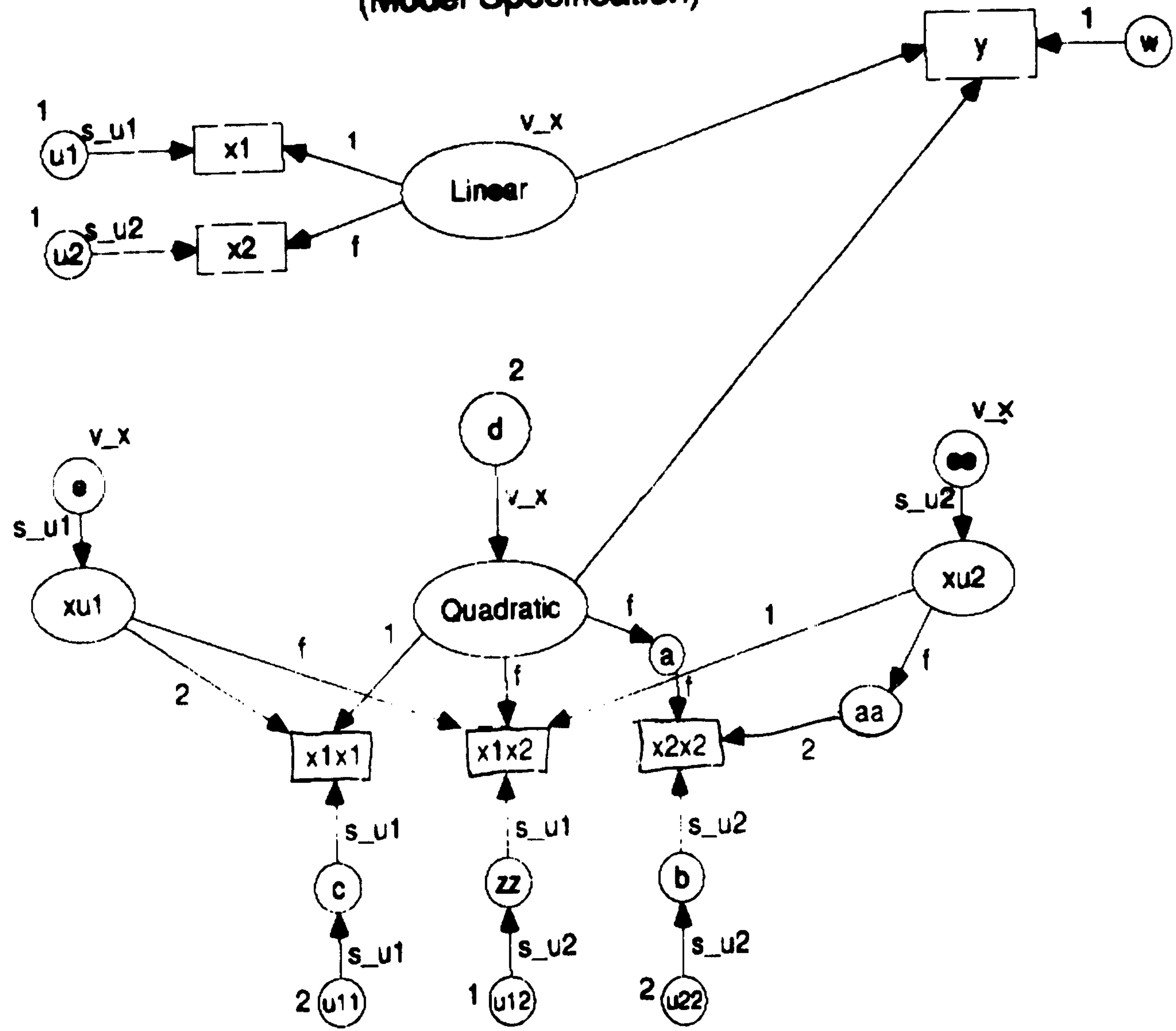
Your model has 9 estimated parameters and 2000 Observed statistics
-2 times log-likelihood of data = 3698.475

AMOS PROGRAM EXAMPLE

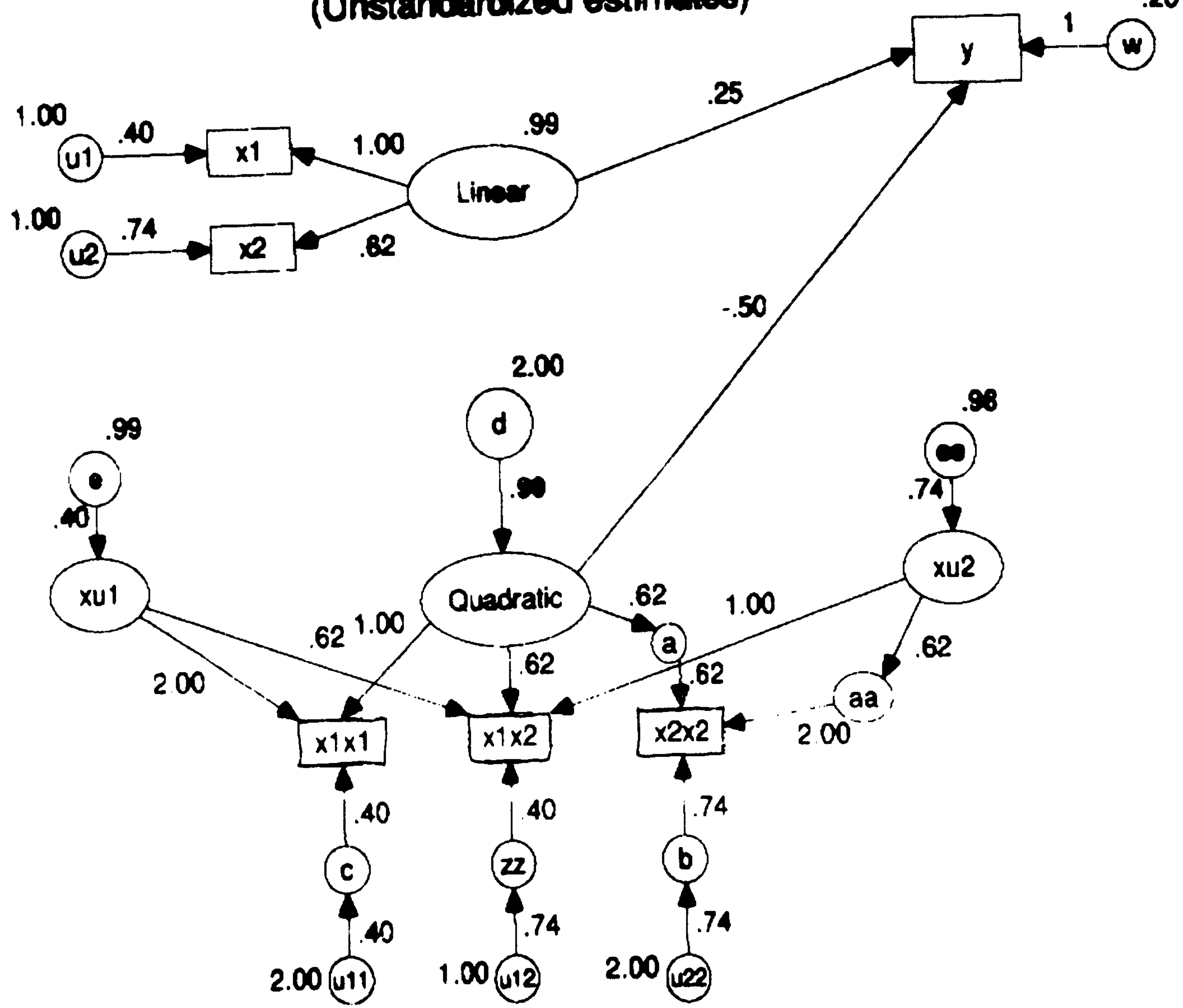
```
! nonlin.amd file contents
! Nonlinear example data from Kenny and Judd, 1984

$Sample size = 500
$input variables
    x1
    x2
    x1x1
    x2x2
    x1x2
    y
$Covariances
 1.150
 .617 .981
 -.068 -.025 2.708
 .075 .159 .729 1.717
 .063 .065 1.459 1.142 1.484
 .256 .166 -1.017 -.340 -.610 .763
```

Note: The nonlin.amw file used in Amos to compute the coefficients is generated from the model specification diagram on the next page. The unstandardized coefficients are then printed for each path (unstandardized estimates) on the diagram (see subsequent page).

Polynomial latent variable regression model
Kenny and Judd (1984), Example 1
Quadratic constraints parameterized according to
Rindskopf (1984, 1985)
(Model Specification)



Polynomial latent variable regression model
Kenny and Judd (1984), Example 1
Quadratic constraints parameterized according to
Rindskopf (1984, 1985)
(Unstandardized estimates)



32

SAS/PROC CALIS EXAMPLE

Interactive models don't often converge and are often difficult to specify. A two-stage fit seems to work best. A short SAS program is provided which will generate the interactive model, given the LINEQS and STD statements of the original measurement model with contraints on the covariances rather than generating nuisance variables. These covariance models run quicker and yield identical results to the model proposed by Kenny & Judd (1984).

I have provided two SAS programs below: (1) the first fits the Kenny & Judd model using LISREL and the appropriate constraints; (2) the second is a small program which the user can employ by putting the LineEQS statements on one card section and the STD statements on another for the measurement model of interest. A punch file is then created which is the Proc CALIS program you need with all of the necessary constraints under a restatement of Kenny & Judd's model with only covariances.

```
*This is the interactive example on page 208 of Loehlin's Latent Variable
Modeling book;
data corrs (type=cov);input x1 x2 z1 z2 x1z1 x1z2 x2z1 x2z2 y _name_ $ 58-61;
_type_='COV';cards;
2.395 . . . . . . . .                    X1
1.254 1.542 . . . . . . .                  X2
.445 .202 2.097 . . . . . .                  Z1
.231 .116 1.141 1.370 . . . . .                  Z2
-.367 -.070 -.148 -.133 5.669 . . . .                X1Z1
-.301 -.041 -.130 -.117 2.868 3.076 . . .              X1Z2
-.081 -.054 .038 .037 2.989 1.346 3.411 . .             X2Z1
-.047 -.045 .039 -.043 1.341 1.392 1.719 1.960 .          X2Z2
-.368 -.179 .402 .282 2.556 1.579 1.623 .971 2.174
Ydata size (type=cov);input x1 x2 z1 z2 x1z1 x1z2 x2z1 x2z2 y _name_ $ 58-61;
cards;
0 0 0 0 0 0 0 0 0                          MEAN
2.395 1.542 2.097 1.370 5.669 3.076 3.411 1.960 2.174    STD
500 500 500 500 500 500 500 500 500                N
;
data corrs (type=cov);set corrs size;
proc print;
proc calis cov data=corrs method=gls maxiter=300 tech=quanew
edf=499;
```

Lineqsx1 = f1 + e1 ,x2 =g (.646) f1 + e2 ,z1 = f2 + e3 ,z2 =h (.685) f2 + e4 ,x1z1 = f1f2 +e1e3 ,x1z2 =h f1f2 +e1e4 ,x2z1 =g f1f2 +e2e3 ,x2z2 =gh f1f2 +e2e4 ,y=c f1 + d f2 + e f1f2 + e5;stde1 =er1 ,e2 =er2 ,e3 =er3 ,e4 =er4 ,e5=er5,e1e3 =er1er3 ,e1e4 =er1er4 ,e2e3 =er2er3 ,e2e4 =er2er4 ,f1 =vf1 ,f2 =vf2 ,f1f2 =vf1f2 ;cove1e3 e1e4 =c1 ,e1e3 e2e3 =c2 ,e1e4 e2e4 =c3 ,e2e3 e2e4 =c4 ,f1 f2 =cf1f2 ;er1er3 =er1 *er3 +vf1 *er3 +vf2 *er1 ;er1er4 =er1 *er4 +vf1 *er4 +h *h *vf2 *er1 ;er2er3 =er2 *er3 +g *g *vf1 *er3 +vf2 *er2 ;er2er4 =er2 *er4 +g *g *vf1 *er4 +h *h *vf2 *er2 ;vf1f2 =vf1 *vf2 +cf1f2 *cf1f2 ;c1 =vf2 *er1 *h ;c2 =vf1 *er3 *g ;c3 =vf1 *er4 *g ;c4 =vf2 *er2 *h ;gh =g *h ;

*Note: all variable names and parameters must be limited to 4 characters,because this program generates product variables and parameters which are products/functions of other variables. 8 character namesparameters will result in generation of illegal SAS names/values;

```
data try;infile cards eof=last;
array name{11} $ xvar1-xvar11;array load{11} $ xload1-xload11;array
eld{11} $ eload1-eload11;array enm{11} $ ename1-ename11;retain xvar1-xvar11 xload1-xload11
eload1-eload11 ename1-ename11 firsty lastx bottom fact1n fact2n;input varname $ symbol $ fload $
fname $ symbol $ eload $ ename $;
 if fname=lag(fname) then cut=0;else cut=1;i+1;
 if cut=1 then do;fact1n=lag(fname);fact2n=fname;lastx=i-1;firsty=i;end;
 name{i}=varname;
 load{i}=fload;
 eld{i}=eload;
 enm{i}=translate(ename,' ',',');
 if eld{i}='1' then eld{i}=' ';
 if load{i}='1' then load{i}=' ';bottom=i;return;last:nvar=i;output;
cards;
 vara = l1 f1 + esd1 e1,varb = l2 f1 + esd2 e2,varc = l3 f1 + esd3 e3,vard = l4 f1 + esd4 e4,varg = m1 f2
 + esd5 e5,varh = m2 f2 + esd6 e6,vari = m3 f2 + esd7 e7,data try2;
infile cards eof=last;array env{11} $ evar1-evar11;
 retain ename1-ename11 evar1-evar11 nvar2;
 input ename $ equals $ evar $ ;I+1;env{i}=translate(evar,'',','); if env{i}='1' then env{i}='';
return;
last:nvar2=i;output;
cards;
  e1 = 1,e2 = 1,e3 = 1,e4 = 1,e5 = 1,e6 = 1,e7 = 1,f1 = 1,f2 = 1,data try;set try;
 drop cut varname symbol fload fname eload ename;data try2;set try2;
 drop I ename equals evar;data try;merge try try2;data try;set try;
 file punch;
 array name{11} $ xvar1-xvar11;
 array load{11} $ xload1-xload11;
 array eld{11} $ evar1-evar11;
 array eld2{11} $ eload1-eload11;
 array enm{11} $ ename1-ename11;
   do i=1 to lastx; do j=firsty to bottom;
 intvar=compress(name{i}||name{j});
 put intvar '=' name{i} '*' name{j} ';';
 end;end;
 put 'Proc Calis cov method=gls;';
 put 'Lineqs';do i=1 to nvar;
 put name{i} '=' load{i} @;
 if i<firsty then put fact1n @;else put fact2n @;
```

```
put '+' @;
if eld2{i}>' ' then put eld2{i} ' ' @;
put enm{i} ',';end;do i=1 to lastx;
do j=firsty to bottom;
intvar=compress(name{i}||name{j});
intload=compress(load{i}||load{j});
intfname=compress(fact1n||fact2n);
intename=compress(enm{i}||enm{j});
put intvar '=' intload intfname '+' intename @;
if i=lastx and j=bottom then put ';';else put ',';
end;end;put 'std';do i=1 to bottom;if eld{i}=' ' then eld{i}='1';
if eld2{i}=' ' then eld2{i}='1';put enm{i} '=' eld{i} ',';end;
do i=1 to lastx;
do j=firsty to bottom;
intename=compress(enm{i}||enm{j});
if eld{i}>'1' then texta=eld{i};else texta=eld2{i};
if eld{j}>'1' then textb=eld{j};else textb=eld2{j};
inteload=compress(texta||textb);
put intename '=' inteload ',';
end;end;
put fact1n '=v' fact1n ',';
put fact2n '=v' fact2n ',';
put intfname '=v' intfname ';';
put 'cov';ind=0;do i=1 to lastx;
do j=firsty to bottom;
intename=compress(enm{i}||enm{j});
do k= i to lastx;
do l=j to bottom;
intenam2=compress(enm{k}||enm{l});
if (i=k or j=l) and not(i=k and j=l) then do;
  ind=ind+1;
  cname='c'||ind;
  cname=compress(cname);
  put intename ' ' intenam2 '=' cname ',';
 end; end; end; end; end; put fact1n ' ' fact2n '=c' intfname ';';
```

***here are the covariance constraints;**

```
do i=1 to lastx;
do j=firsty to bottom;
if eld{i}>'1' then texta=eld{i};else texta=eld2{i};
if eld{j}>'1' then textb=eld{j};else textb=eld2{j};
intename=compress(texta||textb);
put intename '=' @;
 if eld{i}>'1' then put eld{i} @;
 if eld2{i}>'1' then put eld2{i} '*' eld2{i} @;
 put '*' @;
 if eld{j}>'1' then put eld{j} @;
 if eld2{j}>'1' then put eld2{j} '*' eld2{j} @;
 put '+' @;
if load{i}>' ' then put load{i} '*' load{i} '*' @;
```

```
put 'v' fact1n '*' @;
 if eld{j}>'1' then put eld{j} '+' @;
 if eld2{j}>'1' then put eld2{j} '*' eld2{j} '+' @;
 if load{j}>' ' then put load{j} '*' load{j} '*' @;
 put 'v' fact2n '*' @;
 if eld{i}>'1' then put eld{i} ';';
 if eld2{i}>'1' then put eld2{i} '*' eld2{i} ';';
 end;end;
 intfname=compress(fact1n||fact2n);
 put 'v' intfname '=v' fact1n '*v' fact2n '+c' intfname '*c' intfname ';';ind=0;
```

***constraints for the covariances;**

```
do i=1 to lastx;
do j=firsty to bottom;if eld{i}>'1' then intename=compress(eld{i}||eld{j});if eld2{i}>'1' then
intename=compress(eld2{i}||eld{j});
 do k= i to lastx;
 do l=j to bottom;
if eld{k}>'1' then intename=compress(eld{k}||eld{l});
if eld2{k}>'1' then intename=compress(eld2{k}||eld{l});
  if (i=k or j=l) and not(i=k and j=l) then
   do;
   ind=ind+1;
   cname='c'||ind;
   cname=compress(cname);
   put cname '=' @;
    if i=k then do;
                put 'v' fact2n '*' @;
                if eld{i}>'1' then put eld{i} @;
                if eld2{i}>'1' then put eld2{i} '*' eld2{i}  @;
                put '*' load{l} ';';
                end;
    if j=l then do;
                put 'v' fact1n '*' @;
                if eld{j}>'1' then put eld{j} @;
                if eld2{j}>'1' then put eld2{j} '*' eld2{j} @;
                put '*' load{k} ';';
                end;
   end; end;  end; end; end;
do i=1 to lastx;
do j=firsty to bottom;
 if load{i}>' ' and load{j}>' ' then do;
 intfname=load{i}||load{j};
 intfname=compress(intfname);
 put intfname '=' load{i} '*' load{j} ';';
 end; end; end;
```