Structural Equation Modeling Using AMOS

An Introduction

Division of
Statistics + Scientific Computation
THE UNIVERSITY OF TEXAS AT AUSTIN

August 2012

Table of Contents

Section 1: Introduction 3
1.1 About this Document/Prerequisites 3
1.2 Accessing AMOS 3
1.3 Documentation 4
1.4 Getting Help with AMOS 4
Section 2: SEM Basics 5
2.1 Overview of Structural Equation Modeling 5
2.2 SEM Nomenclature 6
2.3 Why SEM? 8
Section 3: SEM Assumptions 8
3.1 A Reasonable Sample Size 8
3.2 Continuously and Normally Distributed Endogenous Variables 9
3.3 Model Identification (Identified Equations) 9
3.4 Complete Data or Appropriate Handling of Incomplete Data 13
3.5 Theoretical Basis for Model Specification and Causality 14
Section 4: Building and Testing a Model using AMOS Graphics 15
4.1 Illustration of the SEM-Multiple Regression Relationship 15
4.2 Drawing a model using AMOS Graphics 19
4.3 Reading Data into AMOS 26
4.4 Selecting AMOS Analysis Options and Running your Model 33
Section 5: Interpreting AMOS Output 35
5.1 Evaluating Global Model Fit 36
5.2 Tests of Absolute Fit 38
5.3 Tests of Relative Fit 38
5.4 Modifying the Model to Obtain Superior Goodness of Fit 39
5.5 Viewing Path Diagram Output 46
5.6 Significance Tests of Individual Parameters 49
Section 6: Putting it all together - A substantive interpretation of the findings 50
References 51

Section 1: Introduction

1.1 About this Document/Prerequisites

This course is a brief introduction and overview of structural equation modeling using the AMOS (Analysis of Moment Structures) software. Structural equation modeling (SEM) encompasses such diverse statistical techniques as path analysis, confirmatory factor analysis, causal modeling with latent variables, and even analysis of variance and multiple linear regression. The course features an introduction to the logic of SEM, the assumptions and required input for SEM analysis, and how to perform SEM analyses using AMOS.

By the end of the course you should be able to fit structural equation models using AMOS. You will also gain an appreciation for the types of research questions well-suited to SEM and an overview of the assumptions underlying SEM methods.

You should already know how to conduct a multiple linear regression analysis using SAS, SPSS, or a similar general statistical software package. You should also understand how to interpret the output from a multiple linear regression analysis. Finally, you should understand basic Microsoft Windows navigation operations: opening files and folders, saving your work, recalling previously saved work, etc.

1.2 Accessing AMOS

You may access AMOS in one of three ways:

1. License a copy from SPSS, Inc. for your own personal computer.
2. AMOS is available to faculty, students, and staff at the University of Texas at Austin via the STATS Windows terminal server. To use the terminal server, you must obtain an ITS computer account (an IF or departmental account) and then validate the account for Windows NT Services. You then download and configure client software that enables your PC, Macintosh, or UNIX workstation to connect to the terminal server. Finally, you connect to the server and launch AMOS by double-clicking on the AMOS program icon located in the STATS terminal server program group. Details on how to obtain an ITS computer account, account use charges, and downloading client software and configuration instructions may be found in http://ssc.utexas.edu/software/stat-apps-server
3. Download the free student version of AMOS from the AMOS development website for your own personal computer. If your models of interest are small, the free demonstration version may be sufficient to meet your needs. For larger models, you will need to purchase your own copy of AMOS or access the ITS shared copy of the software through the campus network. The latter option is typically more cost effective, particularly if you decide to access the other software programs available on the server (e.g., SAS, SPSS, HLM, Mplus, etc.).

1.3 Documentation

The AMOS manual is the *AMOS 16.0 User's Guide* by James Arbuckle and can be found online. It contains over twenty examples that map to models typically fitted by many investigators. These same examples, including sample data, are included with the student and commercial versions of AMOS, so you can easily fit and modify the models described in the AMOS manual.

The previous AMOS manual is the *AMOS 4.0 User's Guide* by James Arbuckle and Werner Wothke; this manual also contains numerous examples. A copy of the *AMOS 4.0 User's Guide* is available at the PCL for check out by faculty, students, and staff at UT Austin. Barbara Byrne has also written a book on using AMOS. The title is *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. The book is published by Lawrence Erlbaum Associates, Inc. Lawrence Erlbaum Associates, Inc. also publishes the journal *Structural Equation Modeling* on a quarterly basis. The journal contains software reviews, empirical articles, and theoretical pieces, as well as a teacher's section and book reviews.

A number of textbooks about SEM are available, ranging from Ken Bollen's encyclopedic reference book to Rick Hoyle's more applied edited volume. Several commonly cited titles are shown below.

Bollen, K.A. (1989). *Structural Equations with Latent Variables.* New York: John Wiley and Sons.

Loehlin, J.C. (1997). *Latent Variable Models.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hoyle, R. (1995). *Structural Equation Modeling: Concepts, Issues, and Applications.* Thousand Oaks, CA: Sage Publications.

Hatcher, L. (1996). *A Step-by-Step Approach to using the SAS System for Factor Analysis and Structural Equation Modeling.* Cary, NC: SAS Institute, Inc.

1.4 Getting Help with AMOS

If you have difficulties accessing AMOS on the STATS Windows terminal server, call the ITS helpdesk at 512-475-9400 or send e-mail to help@its.utexas.edu.

If you are able to log in to the Windows NT terminal server and run AMOS, but have questions about how to use AMOS or interpret output, schedule an appointment with a statistical consultant at SSC statistical consulting or send e-mail to stats@ssc.utexas.edu. **Important note: Both services are available to University of Texas faculty, students, and staff only**. See our Web site at http://ssc.utexas.edu/consulting/ for more details about consulting services, as well as frequently asked questions and answers about EFA, CFA/SEM, AMOS, and other topics.

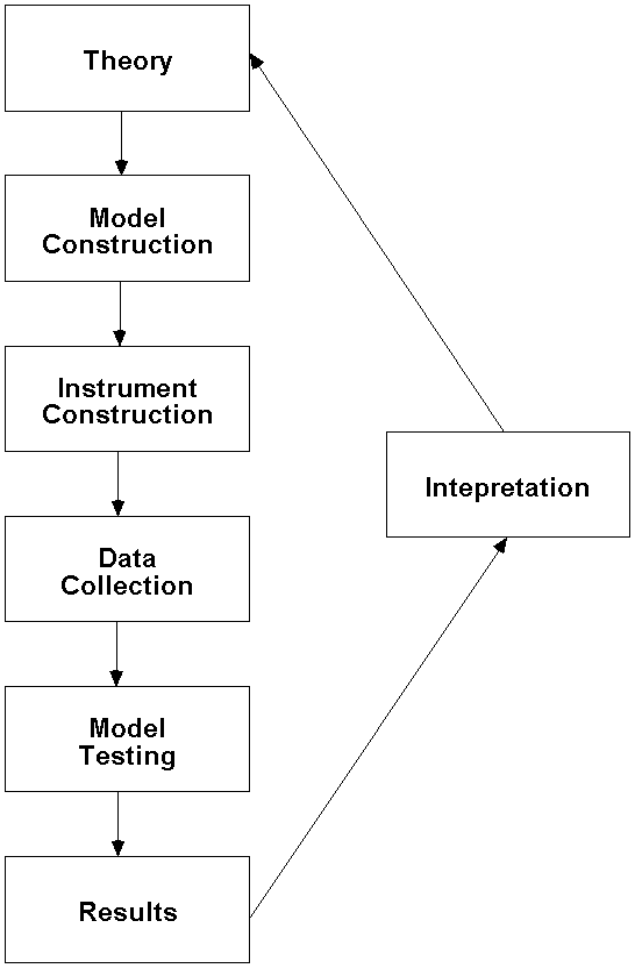
Non-UT and UT AMOS users will find Ed Rigdon's SEM FAQ Web site to be a useful resource; see the information on the SEMNET online discussion group for information on how to subscribe to this forum to post questions and learn more about SEM.

Section 2: SEM Basics

2.1 Overview of Structural Equation Modeling

SEM is an extension of the general linear model (GLM) that enables a researcher to test a set of regression equations simultaneously. SEM software can test traditional models, but it also permits examination of more complex relationships and models, such as confirmatory factor analysis and time series analyses.

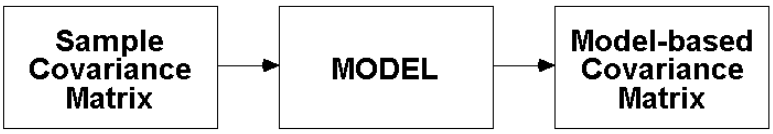
The basic approach to performing a SEM analysis is as follows:



The researcher first specifies a model based on theory, then determines how to measure constructs, collects data, and then inputs the data into the SEM software package. The package fits the data to the specified model and produces the results, which include overall model fit statistics and parameter estimates.

The input to the analysis is usually a covariance matrix of measured variables such as survey item scores, though sometimes matrices of correlations or matrices of covariances and means are used. In practice, the data analyst usually supplies SEM programs with raw data, and the programs convert these data into covariances and means for its own use.

The model consists of a set of relationships among the measured variables. These relationships are then expressed as restrictions on the total set of possible relationships.



The results feature overall indexes of model fit as well as parameter estimates, standard errors, and test statistics for each free parameter in the model.

2.2 SEM Nomenclature

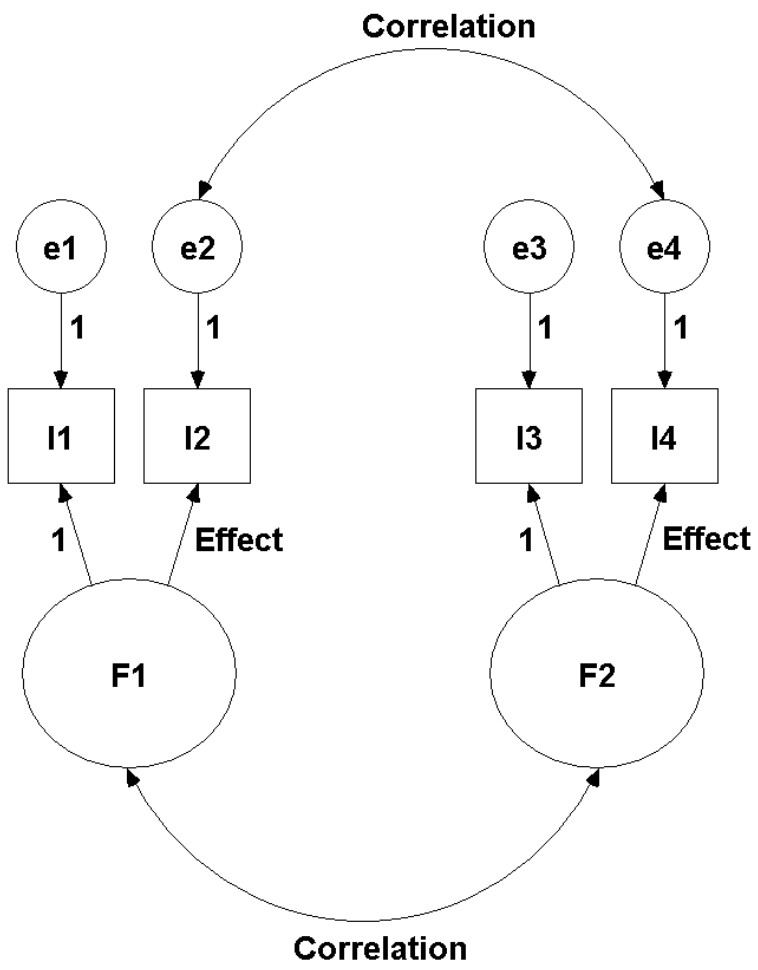
SEM has a language all its own. Statistical methods in general have this property, but SEM users and creators seem to have elevated specialized language to a new level.

Independent variables, which are assumed to be measured without error, are called *exogenous* or *upstream* variables; dependent or mediating variables are called *endogenous* or *downstream* variables.

Manifest or *observed* variables are directly measured by researchers, while *latent* or *unobserved* variables are not directly measured but are inferred by the relationships or correlations among measured variables in the analysis. This statistical estimation is accomplished in much the same way that an exploratory factor analysis infers the presence of latent factors from shared variance among observed variables.

SEM users represent relationships among observed and unobserved variables using path diagrams. Ovals or circles represent latent variables, while rectangles or squares represent measured variables. Residuals are always unobserved, so they are represented by ovals or circles.

In the diagram shown below, correlations and covariances are represented by bidirectional arrows, which represent relationships without an explicitly defined causal direction. For instance, *F1* and *F2* are related or associated, but no claim is made about *F1* causing *F2*, or vice versa.



By contrast, we do claim that *F1* causes the scores observed on the measured variables *I1* and *I2*. Causal effects are represented by single-headed arrows in the path diagram. *F1* and *F2* can be conceptualized as the variance the two indicators share (i.e., what the two indicators have in common.) As you have probably guessed by now, *F1* and *F2* are latent factors; *I1* through *I4* are observed variables. Perhaps they are survey items. *E1* through *E4* are residual or error variances that also cause response variation in *I1* through *I4*. This diagram tells us that scores or responses on survey items one through four are caused by two correlated factors, along with variance that is unique to each item. Some of that unique variance might be due to measurement error.

Some of the paths shown in the diagram are labeled with the number "1". This means that those paths' coefficients have fixed values set to 1.00. These fixed values are included by necessity: they set the scale of measurement for the latent factors and residuals. Alternatively, you can set

the variances of the factors to 1.00 to obtain implicitly standardized solutions. Note: you should not use this latter method when you perform a multiple group analysis.

2.3 Why SEM?

Why would a researcher want to use SEM and have to deal with its own language and, as you shall soon see, some fairly stringent statistical assumptions? SEM has a number of attractive virtues:

- Assumptions underlying the statistical analyses are clear and testable, giving the investigator full control and potentially furthering understanding of the analyses.
- Graphical interface software boosts creativity and facilitates rapid model debugging (a feature limited to selected SEM software packages).
- SEM programs provide overall tests of model fit and individual parameter estimate tests simultaneously.
- Regression coefficients, means, and variances may be compared simultaneously, even across multiple between-subjects groups.
- Measurement and confirmatory factor analysis models can be used to purge errors, making estimated relationships among latent variables less contaminated by measurement error.
- Ability to fit non-standard models, including flexible handling of longitudinal data, databases with autocorrelated error structures (time series analysis), and databases with non-normally distributed variables and incomplete data.
- This last feature of SEM is its most attractive quality. SEM provides a unifying framework under which numerous linear models may be fit using flexible, powerful software.

Section 3: SEM Assumptions

3.1 A Reasonable Sample Size

Structural equation modeling is a flexible and powerful extension of the general linear model. Like any statistical method, it features a number of assumptions. These assumptions should be met or at least approximated to ensure trustworthy results.

According to James Stevens' *Applied Multivariate Statistics for the Social Sciences*, a good general rule for sample size is 15 cases per predictor in a standard ordinary least squares multiple regression analysis. Since SEM is closely related to multiple regression in some respects, 15 cases per measured variable in SEM is not unreasonable. Bentler and Chou (1987) note that researchers may go as low as five cases per parameter estimate in SEM analysis, but only if the

data are perfectly well-behaved (i.e., normally distributed, no missing data or outlying cases, etc.). Notice that Bentler and Chou mention five cases *per parameter estimate* rather than per measured variable. Measured variables typically have at least one path coefficient associated with another variable in the analysis, plus a residual term or variance estimate, so it is important to recognize that the Bentler and Chou and Stevens recommendations dovetail at approximately 15 cases per measured variable, minimum. More generally, Loehlin (1992) reports the results of Monte Carlo simulation studies using confirmatory factor analysis models. After reviewing the literature, he concludes that for this class of model with two to four factors, the investigator should plan on collecting at least 100 cases, with 200 being better (if possible). Consequences of using smaller samples include more convergence failures (the software cannot reach a satisfactory solution), improper solutions (including negative error variance estimates for measured variables), and lowered accuracy of parameter estimates and, in particular, standard errors – SEM program standard errors are computed under the assumption of large sample sizes.

When data are not normally distributed or are otherwise flawed in some way (almost always the case), larger samples are required. It is difficult to make absolute recommendations as to what sample sizes are required when data are skewed, kurtotic, incomplete, or otherwise less than perfect. The general recommendation is thus to obtain more data whenever possible.

3.2 Continuously and Normally Distributed Endogenous Variables

SEM programs assume that dependent and mediating variables (so-called *endogenous* or *downstream* variables in SEM parlance) are continuously distributed, with normally distributed residuals. In fact, residuals from a SEM analysis are not only expected to be univariate normally distributed, their joint distribution is expected to be joint multivariate normal (JMVN) as well. However, this assumption is never completely met in practice.

SEM specialists have developed a number of methods to deal with non-normally distributed variables. These methods are designed for variables that are assumed to have an underlying continuous distribution. For instance, perhaps you administered a Likert scale of self-esteem items to research participants. The scale points tap into points along a continuum of self-esteem, and even though the item data are not continuously distributed, the underlying self-esteem distribution is continuous.

By contrast, other outcome variables are not continuously distributed. For instance, did a patient in a medical study live or die after treatment? Most SEM programs cannot handle these types of nominal-level dependent variables at this time.

3.3 Model Identification (Identified Equations)

As you will soon see, SEM programs require an adequate number of known correlations or covariances as inputs in order to generate a sensible set of results. An additional requirement is that each equation be properly *identified*. Identification refers to the idea that there is at least one unique solution for each parameter estimate in a SEM model. Models in which there is only one possible solution for each parameter estimate are said to be *just-identified*. Models for which

there are an infinite number of possible parameter estimate values are said to be *underidentified*. Finally, models that have more than one possible solution (but one best or optimal solution) for each parameter estimate are considered *overidentified*.

The following equation, drawn from Rigdon (1997) may help make this clearer:

$x + 2y = 7$

In the above equation, there are an infinite number of solutions for x and y (e.g., $x = 5$ and $y = 1$, or $x = 3$ and $y = 2$, or $x = 1$ and $y = 3$, etc.). These values are therefore underidentified because there are fewer "knowns" than "unknowns." A just-identified model is one in which there are as many knowns as unknowns.

$x + 2y = 7$
$3x - y = 7$

For this equation, there are just as many knowns as unknowns, and thus there is one best pair of values ($x = 3$, $y = 2$).

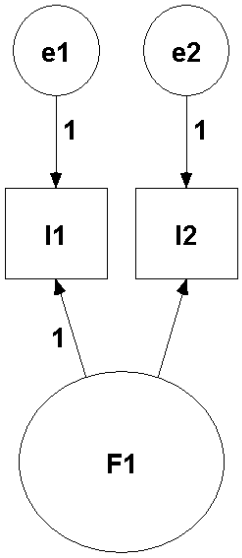
An overidentified model occurs when every parameter is identified and at least one parameter is overidentified (i.e., it can be solved for in more than way--instead of solving for this parameter with one equation, more than one equation will generate this parameter estimate). Typically, most people who use structural equation modeling prefer to work with models that are overidentified. An overidentified model has positive degrees of freedom and may not fit as well as a model that is just identified. Imposing restrictions on the model when you have an overidentified model provides you with a test of your hypotheses, which can then be evaluated using the chi-square statistic of absolute model fit and various descriptive model fit indices. The positive degrees of freedom associated with an overidentified model allows the model to be falsified with the chi-square test. When an overidentified model does fit well, then the researcher typically considers the model to be an adequate fit for the data.

Identification is a structural or mathematical requirement in order for the SEM analysis to take place. A number of rules can be used to assess the identification level of your models, but these rules are not perfect, and they are very difficult (almost impossible, in fact) to evaluate by hand, especially for complex models. SEM software programs such as AMOS perform identification checks as part of the model fitting process. They usually provide reasonable warnings about underidentification conditions.

An additional complication that can arise is *empirical underidentification.* Empirical underidentification occurs when a parameter estimate that establishes model identification has a very small (close to zero) estimate. When the SEM program performs its matrix inversion, that parameter estimate may drop from the solution space defined by the list of model parameters, and the program thus suddenly detects what it perceives to be a structural underidentification problem. Due to the iterative nature of SEM estimation, a parameter estimate such as a variance may start out with a positive value and gradually approach zero with each successive iteration. For example, a path coefficient whose value is estimated as being close to zero may be treated as zero by the SEM program's matrix inversion algorithm. If that path coefficient is necessary to identify the model, the model thus becomes underidentified.

The remedy for all forms of underidentification is to try to locate the source of the identification problem and determine if the source is empirical underidentification or structural underidentification. For structural underidentification, the only remedy is to respecify the model. Empirical underidentification may be correctable by collecting more data or respecifying the model.

An example from Rigdon (1997) may be informative to highlight these issues. Consider the following model:



It contains one factor, *F1*, two error variances or residuals, *e1* and *e2*, and one factor loading value connecting *F1* to *v2*. This model requires four parameters to be estimated: the factor's variance, the two error variances, and the one factor loading.

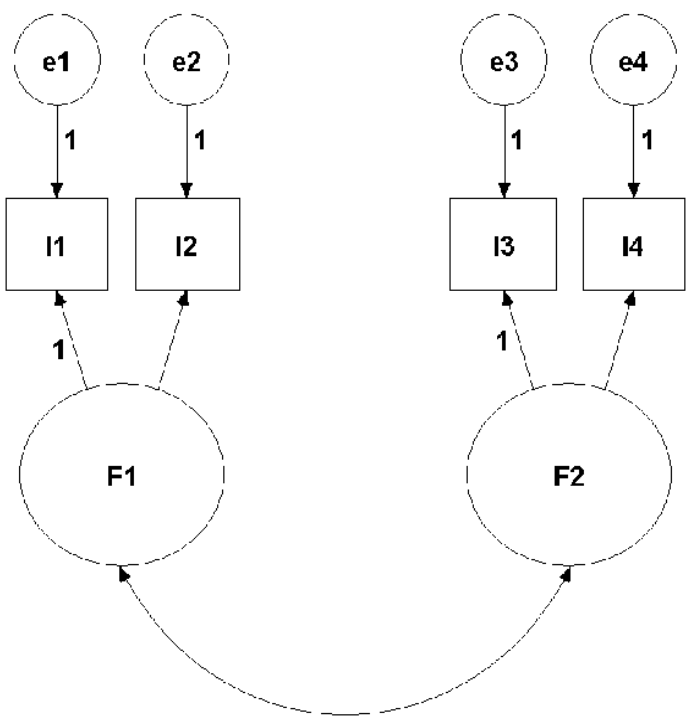
How many of the available inputs can be used in the analysis? Three. How do you know there are three available inputs? You can use the formula

$$[Q(Q + 1)] / 2$$

where *Q* represents the number of measured variables in the database that are used in the model. In this model there are two observed variables, *I1* and *I2*, so via the formula shown above, [2(2+1)]/2 = 3. There are two variances, one for each of the two variables, and one covariance between *I1* and *I2*.

How is it possible to estimate four unknown parameters from three inputs? The answer is that it is not possible: There are three available knowns or degrees of freedom available, but there are four unknown parameters to estimate, so overall, the model has 3 – 4 = -1 degrees of freedom, a clearly impossible state of affairs. This model is clearly underidentified – additional constraints will need to be imposed on this model in order to achieve a satisfactory level of identification.

Now consider a second model:



This new model has [4(4+1)] / 2 = 10 available degrees of freedom because there are four observed variables used in the model. Subtracting four error variances, two factor loadings, and two factor variances, and one covariance between the factors from the 10 available degrees of freedom results in one left over or available degree of freedom. This model is structurally identified. In fact, it is overidentified because there is one positive degree of freedom present.

As it turns out, if the parameter estimate of the covariance between *F1* and *F2* becomes zero or very close to zero, the model can become empirically underidentified because even though it is structurally identified by the covariance specified between *F1* and *F2*, it is not identified on an empirical basis from the computer software's perspective.

In practice, all successfully fitted models are just-identified or overidentified. Typically you want to use overidentified models because these models allow you to test statistical hypotheses, including global model fit (Loehlin, 1992).

3.4 Complete Data or Appropriate Handling of Incomplete Data

Many SEM software programs accept correlation or covariance matrix input. That is, you could compute these matrices yourself using another software package (such as SPSS) and then input them into AMOS or another SEM package for analysis. This feature is useful if you plan to re-analyze a covariance matrix reported in a journal article, for instance.

Usually, however, the preferred mode of analysis uses raw data input: the researcher passes a database, perhaps in SPSS or some other common format (like Microsoft Excel), to the SEM program and it computes the covariances as part of its analysis. What do these programs do about databases without complete data? Typical ad hoc solutions to missing data problems include *listwise* deletion of cases, where an entire case's record is deleted if the case has one or more missing data points, and *pairwise* data deletion, where bivariate correlations are computed only on cases with available data. Pairwise deletion results in different N's for each bivariate covariance or correlation in the database. Another typically used ad hoc missing data handling technique is substitution of the variable's mean for the missing data points on that variable.

But none of these ad hoc missing data handling methods are appealing from a statistical point of view. Listwise deletion can result in a substantial loss of power, particularly if many cases each have a few data points missing on a variety of variables, not to mention limiting statistical inference to individuals who complete all measures in the database. Pairwise deletion is marginally better, but the consequences of using different *n*s for each covariance or correlation can have profound consequences for model fitting efforts, including impossible solutions in some instances. Finally, mean substitution will shrink the variances of the variables where mean substitution took place, which is not desirable. The most important problem with these methods is that they assume that the missing data are missing completely at random, which is often not the case (Little & Rubin, 1987).

What is a researcher to do about missing data? If the proportion of cases with missing data is small, say five percent or less, listwise deletion may be acceptable (Roth, 1994). Of course, if the

five percent (or fewer) cases are not missing completely at random, inconsistent parameter estimates can result. Otherwise, missing data experts (e.g., Little and Rubin, 1987) recommend using a maximum likelihood estimation method for analysis, a method that makes use of all available data points. AMOS features maximum likelihood estimation in the presence of missing data.

3.5 Theoretical Basis for Model Specification and Causality

SEM models can never be accepted; they can only fail to be rejected. This leads researchers to *provisionally accept* a given model. SEM researchers recognize that in most instances there are equivalent models that fit equally as well as their own provisionally accepted model. Any of these models may be "correct" because they fit the data as well as the preferred model. Researchers do their best to eliminate alternative models, and by extension alternative explanations, but this is not always possible. The use of SEM thus entails some uncertainty, particularly with cross-sectional data that are not collected under controlled conditions. (This is also true of other commonly used models such as ANOVA and multiple regression techniques.)

For this reason, SEM software programs require researchers to be very explicit in specifying models. While models that fit the data well can only be provisionally accepted, models that do not fit the data well can be absolutely rejected. For instance, if you fit a single factor confirmatory factor analysis model to a set of ten survey items, and the model is rejected, you can be confident that a single factor is not sufficient to explain the items' shared variance, a useful finding, particularly if you believe that one common factor is not enough to explain the items' shared variance. Suppose you ran a single factor model and then a dual factor model on the same set of ten items; the former model is rejected but the latter model is not rejected. Now you know that more than one factor is needed to account for the shared variance among the measured items.

In addition to evaluating the absolute goodness of fit of single models, you can also evaluate competing models by using likelihood ratio chi-square tests to compare them. Returning to the previous example, you could compare the single and dual factor models to each other using a statistical test. If that test statistic is significant, you can conclude that the more complex two factor model fits the data better than the one factor model. On the other hand, had you found no significant difference between the two models, you could conclude that the one factor model fit the data just as well as the two factor model.

Not all models are directly comparable in this way. Only models where you can derive a simpler model by imposing a set of constraints or restrictions onto a more complex model may be compared in this manner. These models are called *nested* models. When you want to compare models that are not directly comparable, you can use various descriptive criteria to compare these *non-nested* models.

Section 4: Building and Testing a Model using AMOS Graphics

4.1 Illustration of the SEM-Multiple Regression Relationship

In essence, SEM is a multivariate extension of the multiple linear regression model with one dependent (Y) variable:

$$y = i + Xb + e$$

where y = a vector containing observed scores on the dependent variable, i is a vector of 1's representing the y-intercept, X is a matrix of continuously distributed or categorical (dummy-coded) independent variables, b is the vector of regression weights, and e represents the vector of residual or error or leftover scoring unexplained by the model.

SEMs consist of series of multiple regression equations – all equations are fitted simultaneously. In fact, you can conduct a multiple regression analysis using SEM software.

A typical multiple linear regression analysis produces several statistics, including an overall test of model fit and tests of individual parameter estimates. In addition, the analysis prints out unstandardized regression coefficients, standard errors for those coefficients, and a standardized version of the regression coefficients. In addition, a squared multiple correlation or R^2 for the regression equation indicates the proportion of variance in the dependent variable accounted for by the set of independent variables in the multiple regression equation. As you will soon see, AMOS produces these very same statistics, though it generates them for multiple equations rather than for a single equation, as is the case in ordinary least-squares regression.

Suppose you have a database with three continuous predictor variables: education level, a socioeconomic indicator, and feelings of powerlessness measured in 1967. There is one continuous dependent variable, feelings of powerlessness measured in 1971. These data are simulated based on the results reported in a larger study by Wheaton, Muthén, Alwin, and Summers (1977).

If you run a multiple regression analysis in SPSS for Windows using these variables, you will obtain the following results (some output is omitted in the interest of conserving space):

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.56[a]	.32	.31	2.62

a. Predictors: (Constant), Powles67, SEI, Educ

Correlations

		Powles71	Educ	SEI	Powles67
Pearson Correlation	Powles71	1.00	-.36	-.33	.53
	Educ	-.36	1.00	.54	-.42
	SEI	-.33	.54	1.00	-.32
	Powles67	.53	-.42	-.32	1.00
Sig. (1-tailed)	Powles71	.	.00	.00	.00
	Educ	.00	.	.00	.00
	SEI	.00	.00	.	.00
	Powles67	.00	.00	.00	.
N	Powles71	932	932	932	932
	Educ	932	932	932	932
	SEI	932	932	932	932
	Powles67	932	932	932	932

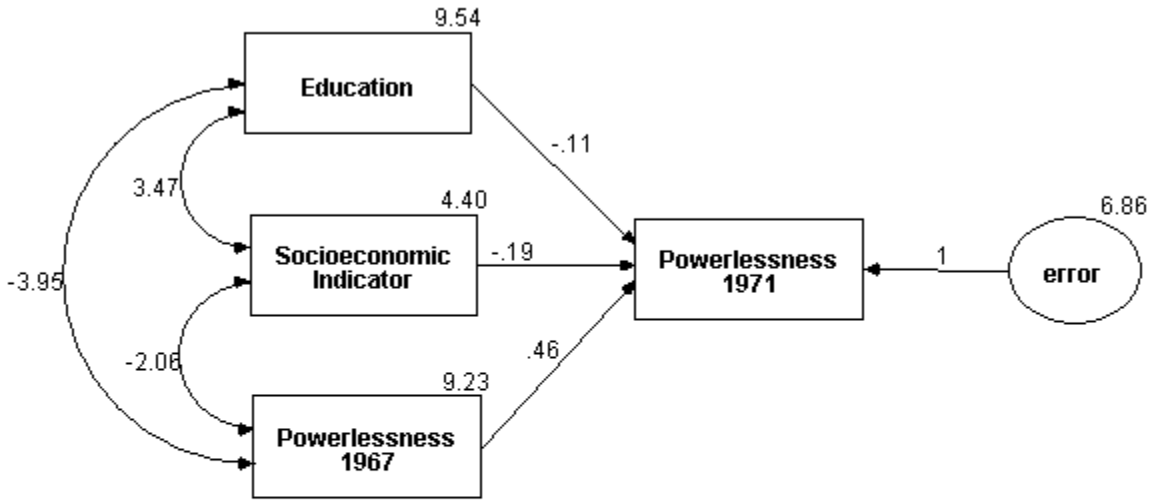
Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	6.11	1.02		6.02	.00
	Educ	-.11	.03	-.11	-3.24	.00
	SEI	-.19	.05	-.13	-3.98	.00
	Powles67	.46	.03	.44	14.61	.00

a. Dependent Variable: Powles71

The salient output from SPSS shows the correlations among each of the predictors, as well as the predictors and the dependent variable. An R^2 value of .32 is then shown, with unstandardized and standardized regression coefficients and significance tests shown in the final table.

Now consider the equivalent model fit in AMOS:



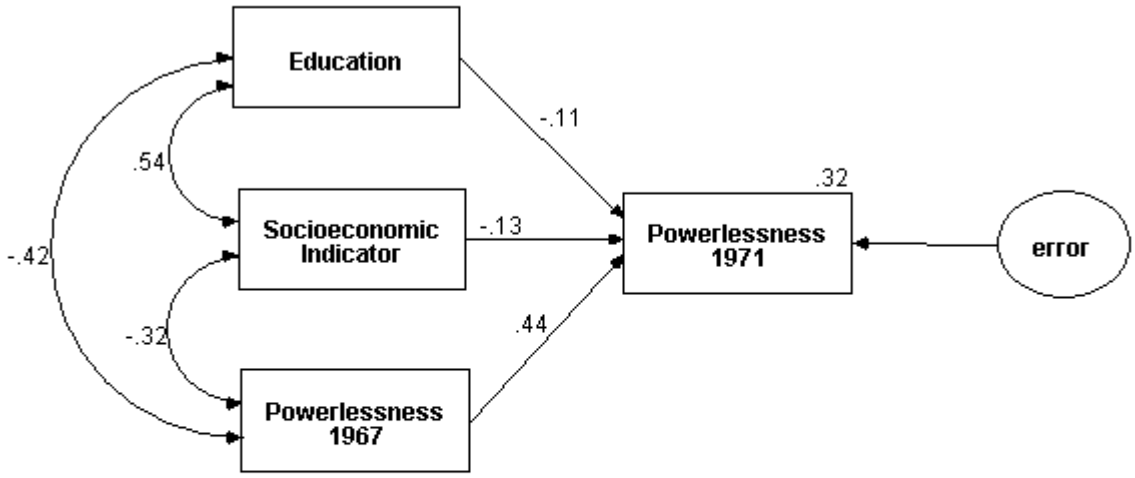
Wheaton Data
Regression Example
Unstandardized estimates

The three predictors are allowed to covary; predictors' covariances are shown in the diagram. For example, the covariance between education and the socioeconomic indicator is 3.47. The variance of each predictor appears above its rectangle. The unstandardized regression weights, which correspond to the *B* coefficients displayed in the SPSS output, link the predictor variables to the dependent variable's rectangle in the AMOS diagram. For instance, the unstandardized regression coefficient of the education variable is –.11. Since the unstandardized regression coefficient represents the amount of change in the dependent variable per single unit change in the predictor variable, this result suggests that for every single unit of increase in education level, powerlessness in 1971 is reduced by -.11 units in the population from which Wheaton et al. drew their sample.

The diagram also features a residual variance associated with the measurement of powerlessness in 1971. Since this variance is not directly measured, it is represented as a latent variable in the AMOS diagram. The estimate of the error variance is 6.86.

In addition to the unstandardized regression coefficients, SPSS also produces standardized regression coefficients and an R^2 value that summarizes the proportion of variance in the

dependent variable explainable by the collective set of the predictors. AMOS produces these statistics as well, but it uses a separate diagram to display the standardized coefficients and R^2 value.



Wheaton Data
Regression Example
Standardized estimates

In this diagram the covariances have been replaced by correlations. For example, the correlation between the socioeconomic indicator and the powerlessness measure in 1967 is -.32, the same value shown in the SPSS Correlations table above. Standardized regression coefficients now link the predictors to the dependent variable, and the R^2 value for the dependent variable appears above its rectangle on the diagram. The standardized regression weights represent the amount of change in the dependent variable that is attributable to a single standard deviation unit's worth of change in the predictor variable.

AMOS also produces tabular output similar to that of SPSS that displays the unstandardized and standardized regression coefficients, the standard error estimates of the unstandardized regression coefficients, and tests of statistical significance of the null hypothesis that each unstandardized regression coefficient equals zero. The AMOS tabular output will be discussed at more length below.

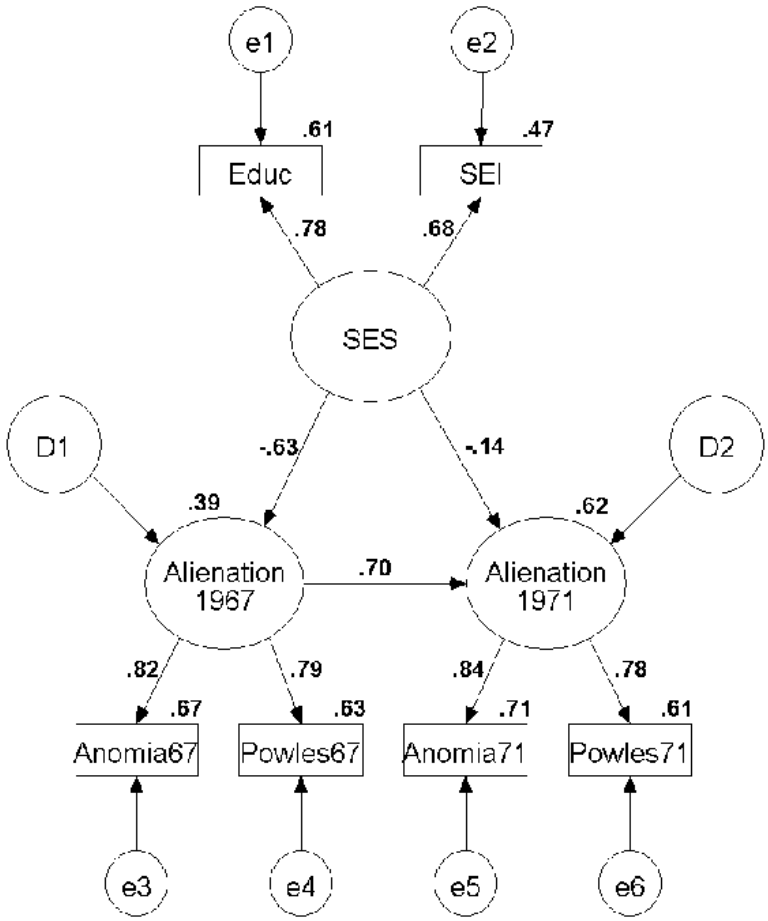
The regression analysis shown above is limited in several ways:

- Multiple dependent or outcome variables are not permitted
- Mediating variables cannot be included in the same single model as predictors
- Each predictor is assumed to be measured without error
- The error or residual variable is the only latent variable permitted in the model
- Multicollinearity among the predictors may hinder result interpretation

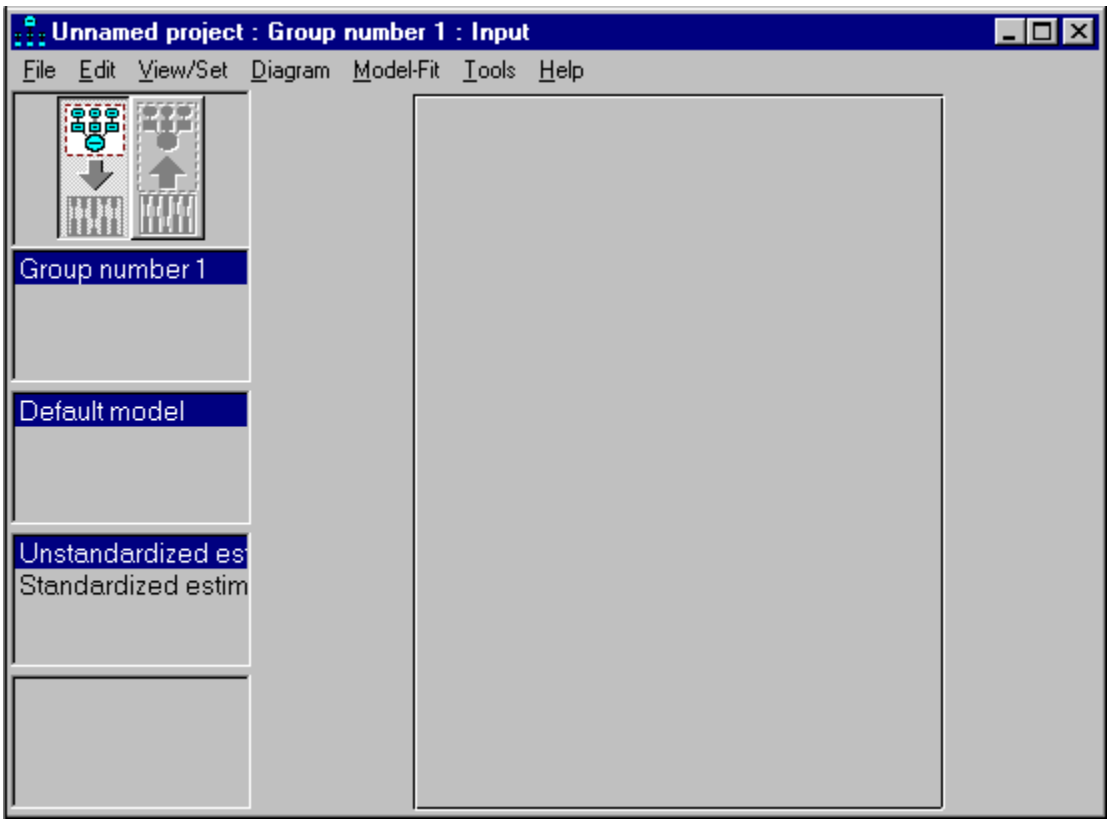
AMOS can fit models that are not subject to these limitations. An example of such a model applied to the Wheaton et al. database appears in the next section.

4.2 Drawing a model using AMOS Graphics

Suppose you want to test a model of the stability of alienation over time, as measured by anomia and powerlessness feelings at two measurement occasions, 1967 and 1971, as well as education level and a socioeconomic index. The path diagram, including standardized regression coefficients, appears below.



How can you test this model and develop from scratch a publication quality path diagram like the one shown above? The first step is to launch AMOS. If you are using the UT server, click on the Start menu from the Windows taskbar, select All Programs, select Statistics, then choose AMOS 16 Graphics. You will now see a blank AMOS Graphics diagram page that looks like this:



You will also see a floating toolbar appear. The toolbar may partially obscure the AMOS drawing area, but you can move it out of the way by dragging it to the side of your computer screen. Select the toolbar by single-clicking on it with your mouse pointer. If you place your mouse pointer over any tool icon with the toolbar selected, AMOS displays the name and function of the tool. You can also right-click on the tool icon for more detailed AMOS help on how to use the tool.

To add or remove individual tools from the toolbar, choose **Move Tools** from the **Tools** menu in AMOS Graphics. You may need to resize your toolbar from time to time to view all available tools, particularly if you have many tools displayed or you have a small monitor.

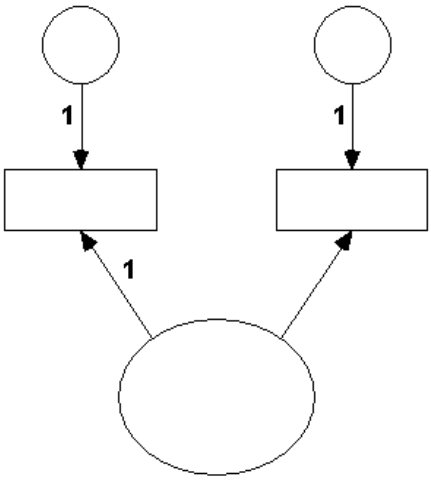
Tools are activated by single-clicking on their icons with your mouse pointer. A tool that is active or in use will have an icon that appears to be depressed or lowered. To deactivate a particular tool, single-click on its icon once again. Its icon should change to look like the other tools on the toolbar.

Select the **Draw Latent Variables and Indicators** tool from the toolbar by first single-clicking on the toolbar to make it active and then single-clicking on the **Draw Latent Variables and Indicators** button to make it active. This tool resembles a factor analysis model with three indicator variables:

Move your mouse pointer to the drawing surface and draw an oval by clicking and holding your mouse button. Once you have a satisfactory oval drawn, click on the oval twice. You will now have a latent variable with two observed indicators.

Perhaps your latent and observed variables are too close to the edge of the AMOS drawing space. To move them closer to the center of the diagram, click on the **Preserve Symmetries** button.

Next, click on the red moving truck tool **(Move Objects)**; then click and drag the latent variable and its indicators closer to the center of the AMOS drawing space. When you are finished, you will have a diagram that looks like this:



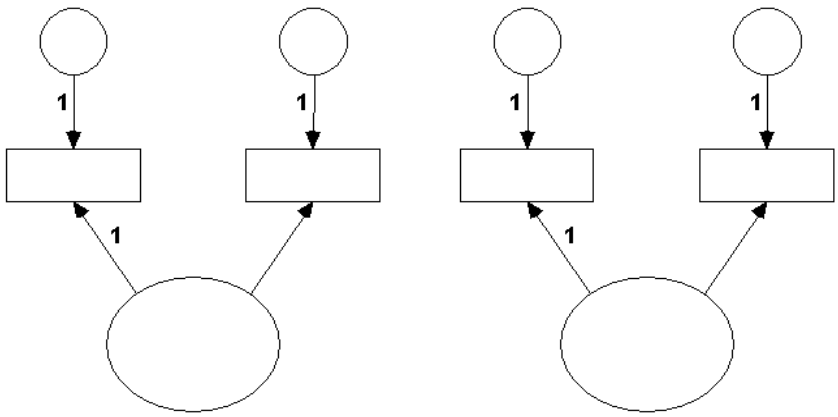
So far, you have drawn part of the desired model. This part of the model could represent a latent variable called *Socioeconomic Status* (*SES*) with two observed indicators, *Education*, and *Socioeconomic Index* (*SEI*). Recall that rectangles represent observed or measured variables, while ovals represent latent or unobserved variables. The two ovals leading to the two observed variables represent the residuals for those observed variables, while the larger oval represents their shared variance. This shared variance is captured by the latent variable, *Socioeconomic Status*.

The model has two more segments that resemble this section (alienation at 1967 and alienation at 1971), so you can copy the portion of the model you have already built. To do that:

- Turn off the **Preserve Symmetries** tool button by clicking on it.
- Click on the **Select All Objects** tool (represented by a raised hand with four fingers extended). The entire diagram should change color from black to blue.
- Click on the **Duplicate Objects** tool icon (it resembles a photocopier), click on the latent variable's oval, and drag your mouse pointer to the right. You should see a replica of the single latent variable-two indicator variable structure appear.

Hint: If you hold down the shift key while you perform this operation, the new portion of the diagram will be horizontally aligned with the previous diagram section.

When you finish with this step, your drawing should look like this:

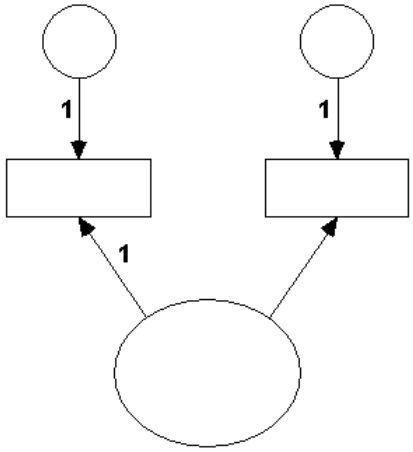


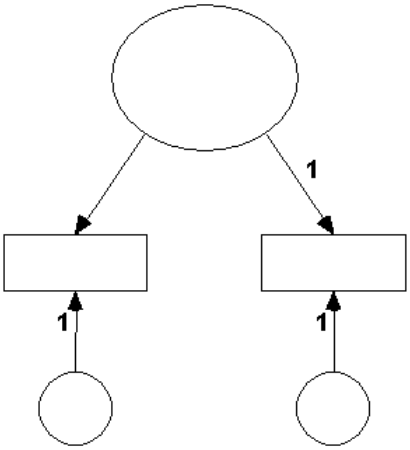
Click on the **Deselect Objects** tool button to remove the object selection. This button resembles a hand with four fingers retracted or bent.

Rotate the indicators of this second latent variable so that they point "down" on the AMOS diagram space instead of "up". Select the **Rotate Indicators** tool button.

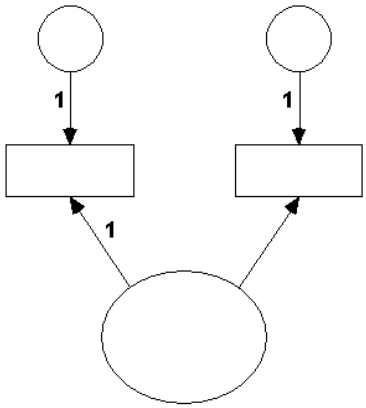
Click once on the second latent variable's oval. The two indicators and their residuals rotate 90 degrees clockwise. Click the oval once more. The indicators rotate another 90 degrees. Click on the **Rotate Indicators** tool button to deactivate it.

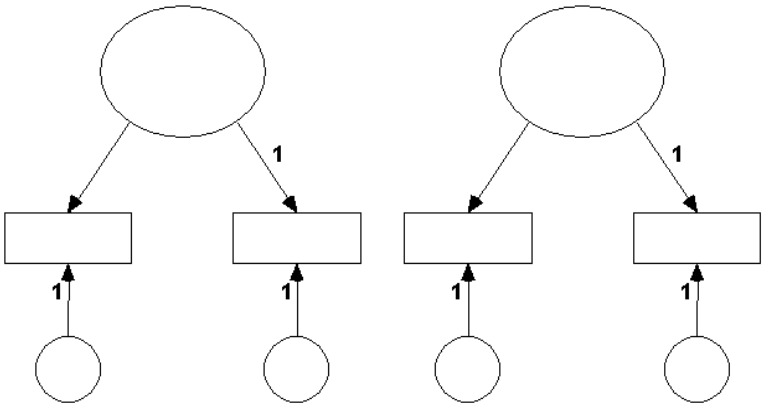
Click on the **Preserve Symmetries** tool and then click on the **Move Objects** tool. Click on the second latent variable and move it beneath the first latent variable and to the left on the diagram. When you finish, your diagram should look like this:





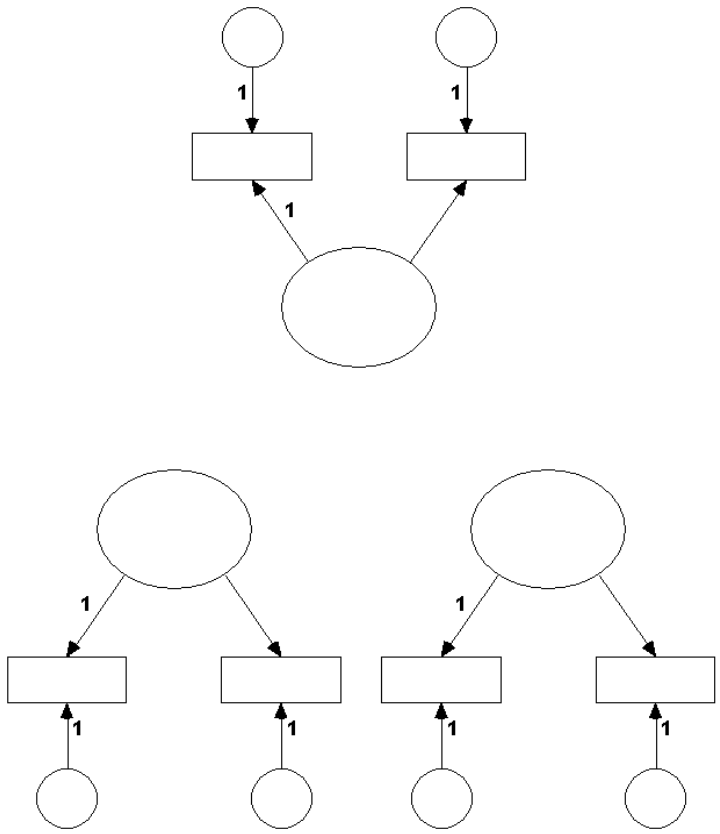
Copy the latent and observed variables in the lower set by selecting the **Duplicate Objects** button from the toolbar once again, selecting the latent variable of the "lower" latent variable, and drag it to the right. Your diagram will now look like this:





Notice that a number of the paths are fixed to a value of 1.00. These are present to ensure proper model identification. Perhaps you want to have those paths be consistently on the left side of each variable set. To do that, use the **Reflect Indicators** tool.

Once you have selected the tool button, click once on each of the two “lower” latent variables. Your updated diagram will now look like this:



Congratulations! You have just specified the measurement part of your model. Now you must tell AMOS where to find the data for conducting the analysis and you must label the observed variables.

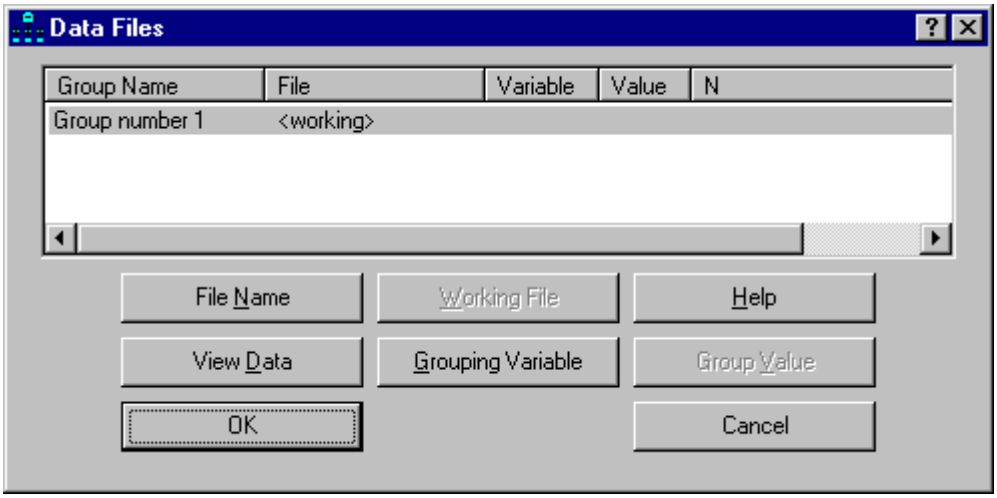
4.3 Reading Data into AMOS

To read data into AMOS, choose **Data Files** from the **File** menu.

File

Data Files

The following dialog box appears:



Click on **File Name** to specify the name of the data file. The **View Data** button launches a data file's external application. For instance, if you specify a SPSS data file using the **File Name** button, and you then click **View Data**, AMOS will launch SPSS so that you can view the contents of the data file. The **Grouping Variable** button allows you to specify a grouping variable within a database, so you can easily set up and test models that involve multiple groups of subjects.

Currently AMOS reads the following data file formats:

- Access
- dBase 3 – 5
- Microsft Excel 3, 4, 5, and 97
- FoxPro 2.0, 2.5 and 2.6
- Lotus wk1, wk3, and wk4
- SPSS *.sav files, versions 7.0.2 through 9.0 (both raw data and matrix formats)
- Comma-delimited text files (semicolon-delimited in countries where the comma is used as a decimal separator).

AMOS recognizes empty or blank cells in MS Excel and Access as missing data. System-missing data (represented by blank cells in the SPSS data editor spreadsheet) are also properly recognized by AMOS as missing data points. Comma-delimited data files should have two consecutive commas listed to represent a missing data point.

Simulated data for this example model are in an SPSS for Windows data file called **Wheaton-generated.sav**. If you have not already downloaded this data file, you can retrieve it at the following Web address:

http://ssc.utexas.edu/software/software-tutorials#amos

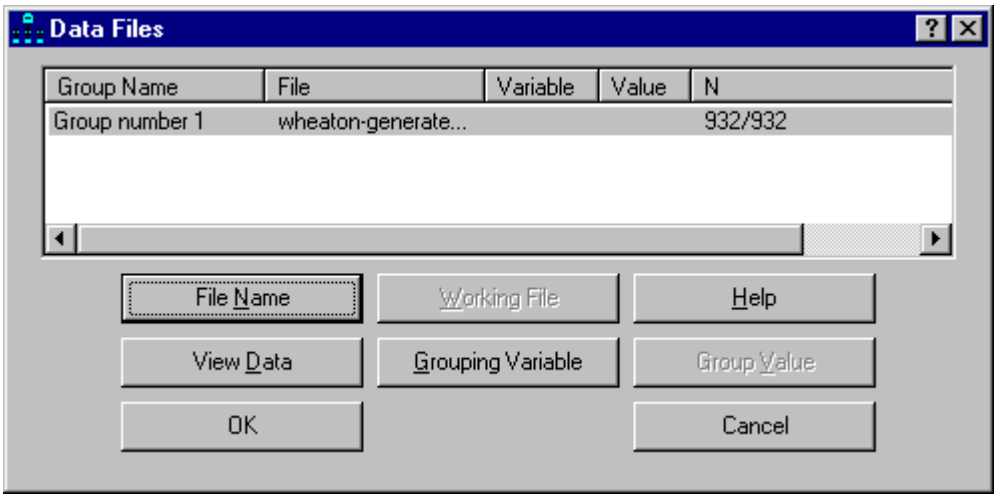
To read these data into AMOS, select **Data Files**… from the AMOS Graphics **File** menu.

File
  Data Files

In the dialog box that appears, click on the **File Name** button. AMOS will produce a file opening dialog box that resembles the box shown below.

In this example, the Wheaton-generated SPSS database has been located and selected in the **File Name** slot of this dialog box. You should follow suit, using your copy of AMOS Graphics: locate the Wheaton-generated.sav SPSS data file on your computer, and choose it using AMOS.

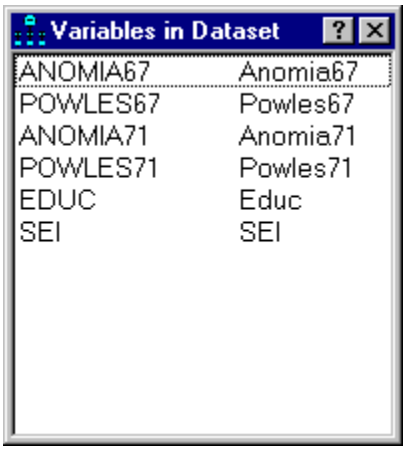
Once you have located and specified the appropriate data file, click on the **Open** button. You will now see the following window:



Click **OK** to return to AMOS Graphics. You have now made the data available to AMOS.

Once AMOS receives the data, you can label observed variables. Select the **Variables in Dataset** option from the **View/Set** menu to see a window like the one shown below.

View/Set
Variables in Dataset

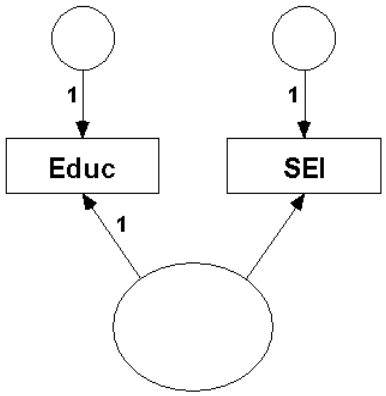


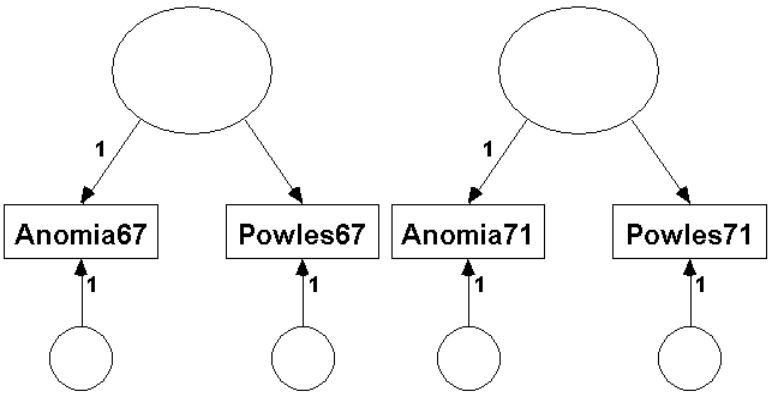
Select each of these entries one at a time and drag them onto the appropriate rectangles on the path diagram. You may notice that some of the variable names are too large to fit inside their rectangles. One partial remedy for this problem is to enlarge the entire model by clicking on the **Resize Diagram to Fit the Page** tool bar icon.

You can also use the **Shape Change** tool icon to alter the shape of the rectangles so that they are wide enough to hold the variable names:

Select each of the rectangles using the **Select Single Objects** tool (it resembles a hand with the index finger extended).

Next, click on the **Shape Change** tool icon to alter the shape of the rectangles. Finally, you can reduce the size of the variable name font by double-clicking on the rectangle for a particular variable. After you make these adjustments, your finished measurement model diagram should look like this:





There are three remaining tasks left. You must:

1. Name the latent variables.
2. Specify the relationships among the latent variables (the structural model).
3. Create residual terms for any latent variables predicted by other variables in the model. Any variable predicted by another variable or set of variables must have a residual value.

To name the upper latent variable, double-click on it and enter its name (SES) in the **Variable Name** text box. Then close the window. To name the lower left latent variable, double-click on it and enter its name, *Alienation 1967*, in the **Variable Name** text box. Repeat this process for the lower right latent variable, *Alienation 1971*.

For large models, there is a **Name Unobserved Variables** macro available under the **Tools** menu. This macro will name the unobserved variables automatically. Select and run this macro now to name the remaining unobserved variables that have not yet been named in the diagram

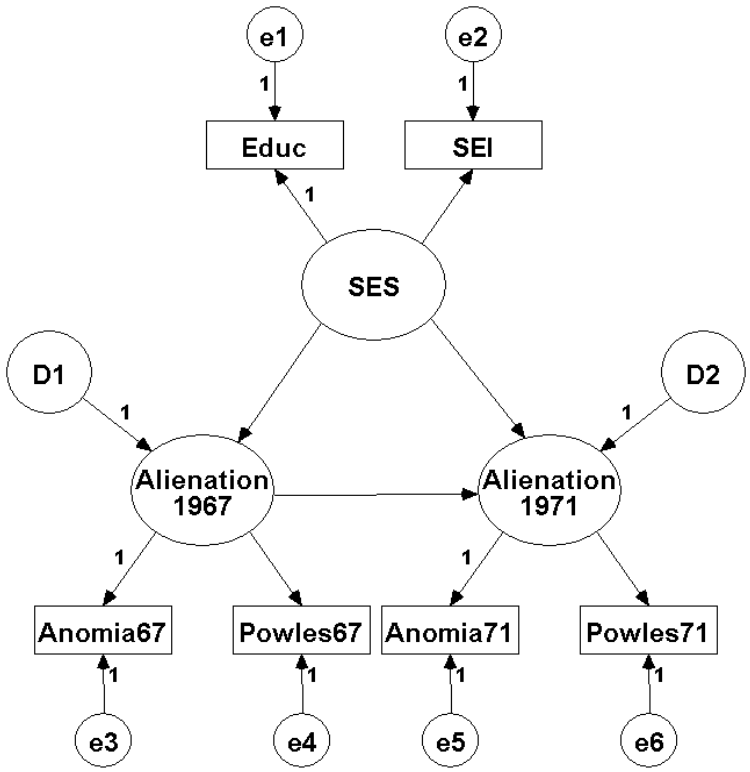
(i.e., the residuals). After the macro runs, each residual will have a name, beginning with *e1*.

To specify the paths among the latent variables, use single-headed arrows to represent causal relationships and dual-headed arrows to represent bidirectional relationships (correlations).

To create a residual for a latent variable, use the **Add Unique Variable** tool icon.

Select the tool's icon on the toolbar. Next, click once on a latent variable to create a unique residual variable for that latent variable. You can click on the latent variable again to move the position of the new residual variable 45 degrees to the right (clockwise). By clicking on the latent variable repeatedly, you can move the new residual variable around the original latent variable in 45 degree increments until you find a location where it looks best on the diagram.

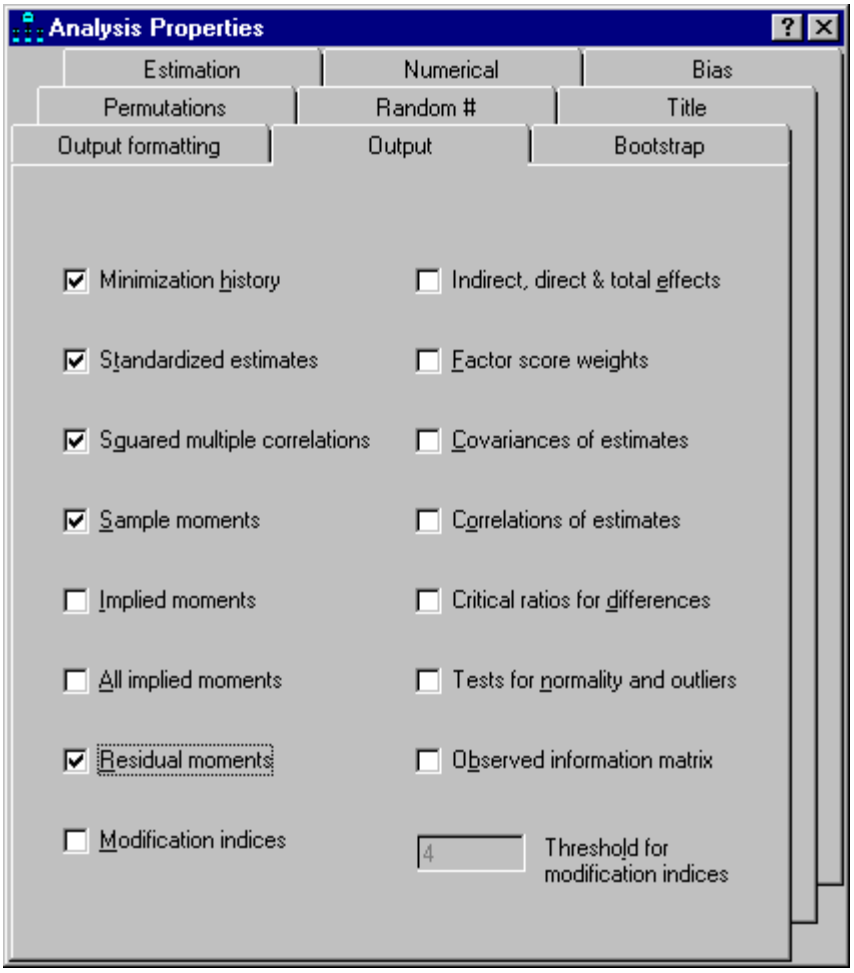
After you make these adjustments, your finished model diagram should resemble the following image.



After you have drawn your model, be sure to save it by choosing the **File** menu, and then selecting **Save As**. When you save the model, AMOS automatically creates two back-up model files in the same directory in which you have your original diagram. The original diagram file will have the extension **.amw**; the back-up files will have the extensions **.bk1** and **.bk2**, respectively.

4.4 Selecting AMOS Analysis Options and Running your Model

You are now ready to run the model. But first, select **Analysis Properties** from the **View/Set** menu. Investigate the various tabs and options available under each tab. One of the more interesting choices is the **Output** tab.



In this analysis, the investigator requests a number of options, including a standardized solution, squared multiple correlations, the sample covariance matrix, and the covariance matrix of the residuals remaining after AMOS fits the model.

Next, examine the **Estimation** tab. Notice that this tab provides a check box that allows you to estimate means and intercepts.

Hint: If your database has any cases with incomplete data, AMOS will require you to estimate means and intercepts; so you must select this check box if your database has any missing data on observed variables included in your model.

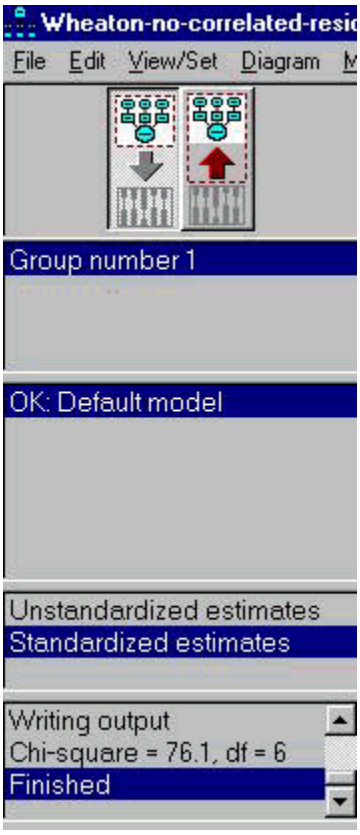
Because this model's database does not contain any missing data and we are not interested in means at present, we leave the **Estimation** tab settings at their default values.

Before you run the model, be sure to save it by choosing **Save As** from the **File** menu and saving a copy of the model file to an appropriate location on your computer's disk drive.

To run the model, close the **Analysis Properties** window and click on the **Calculate Estimates** tool icon. It resembles an abacus:

Section 5: Interpreting AMOS Output

You can check to see if your model ran successfully by examining the left-hand side of the AMOS Graphics drawing area. Consider the following screen shot taken after the Wheaton model finished running:

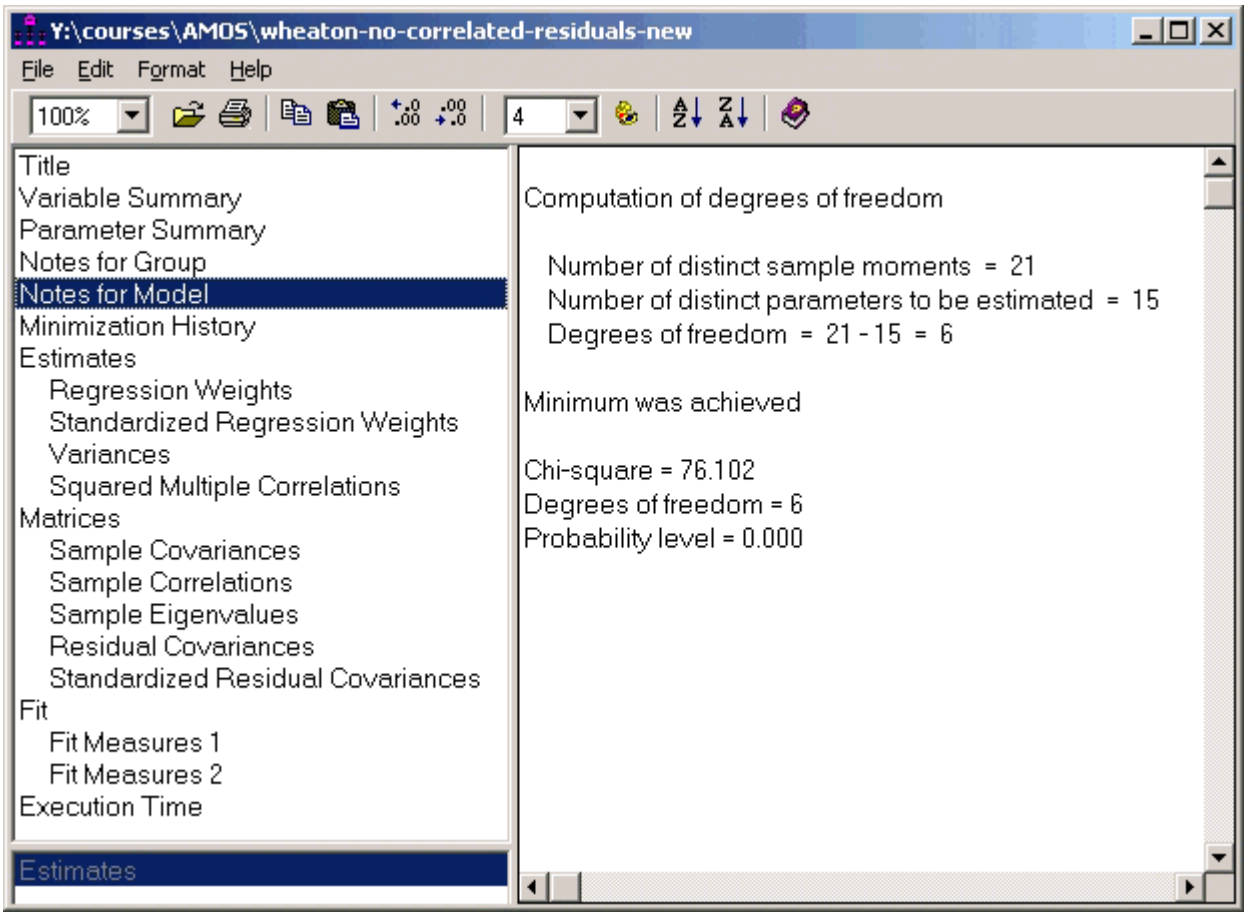


The bottom portion of this output shows a chi-square value of 76.10 with 6 degrees of freedom. This test statistic tests the overall fit of the model to the data. The null hypothesis under test is that the model fits the data, so you hope to find a small, non-significant chi-square value for this test. The significance test for the chi-square model fit statistic is described in the next section.

5.1 Evaluating Global Model Fit

So far, none of the output you have seen contains tests of statistical significance. AMOS produces such tests. They are stored in a spreadsheet table file. You can access this output by selecting the View Table Output tool.

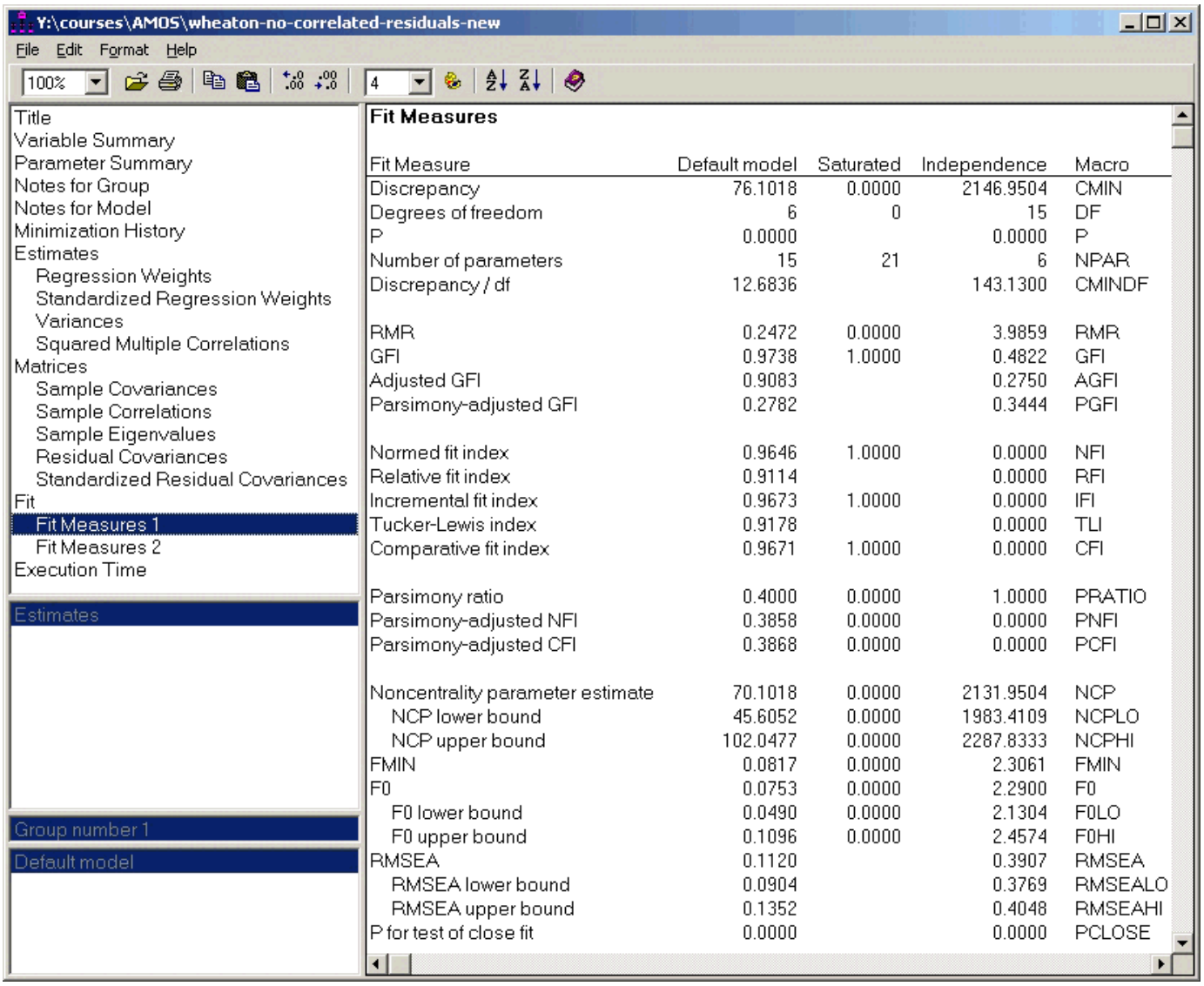
The table that appears contains a list of output categories on the left side of the window and an open space on the right. When you select a particular subset of output for display, it appears in the right-hand space. Consider the *Notes for Model* section of the output.



You should always examine the *Notes for Model* section of the AMOS output after each AMOS analysis finishes because AMOS will display most errors and warnings in this section of the

output. In the output shown above, AMOS reports that the minimum was achieved with no errors or warnings. The chi-square test of absolute model fit is reported, along with its degrees of freedom and probability value. The interpretation of these results will be discussed in further detail below.

The absence of errors or warnings in this section of the output means that it is safe for you to proceed to the next output section of interest, the *Fit Measures* output.



Fit Measures

Fit Measure	Default model	Saturated	Independence	Macro
Discrepancy	76.1018	0.0000	2146.9504	CMIN
Degrees of freedom	6	0	15	DF
P	0.0000		0.0000	P
Number of parameters	15	21	6	NPAR
Discrepancy / df	12.6836		143.1300	CMINDF
RMR	0.2472	0.0000	3.9859	RMR
GFI	0.9738	1.0000	0.4822	GFI
Adjusted GFI	0.9083		0.2750	AGFI
Parsimony-adjusted GFI	0.2782		0.3444	PGFI
Normed fit index	0.9646	1.0000	0.0000	NFI
Relative fit index	0.9114		0.0000	RFI
Incremental fit index	0.9673	1.0000	0.0000	IFI
Tucker-Lewis index	0.9178		0.0000	TLI
Comparative fit index	0.9671	1.0000	0.0000	CFI
Parsimony ratio	0.4000	0.0000	1.0000	PRATIO
Parsimony-adjusted NFI	0.3858	0.0000	0.0000	PNFI
Parsimony-adjusted CFI	0.3868	0.0000	0.0000	PCFI
Noncentrality parameter estimate	70.1018	0.0000	2131.9504	NCP
NCP lower bound	45.6052	0.0000	1983.4109	NCPLO
NCP upper bound	102.0477	0.0000	2287.8333	NCPHI
FMIN	0.0817	0.0000	2.3061	FMIN
F0	0.0753	0.0000	2.2900	F0
F0 lower bound	0.0490	0.0000	2.1304	F0LO
F0 upper bound	0.1096	0.0000	2.4574	F0HI
RMSEA	0.1120		0.3907	RMSEA
RMSEA lower bound	0.0904		0.3769	RMSEALO
RMSEA upper bound	0.1352		0.4048	RMSEAHI
P for test of close fit	0.0000		0.0000	PCLOSE

The Fit Measures output contains five columns. The first column shows the name of each fit measure. The second column, labeled *Default model*, contains the fit statistics for the model you specified in your AMOS Graphics diagram.

The next two columns, labeled *Saturated* and *Independence*, refer to two baseline or comparison models automatically fitted by AMOS as part of every analysis. The *Saturated* model contains as

many parameter estimates as there are available degrees of freedom or inputs into the analysis. The *Saturated* model is thus the least restricted model possible that can be fit by AMOS. By contrast, the *Independence* model is one of the most restrictive models that can be fit: it contains estimates of the variances of the observed variables only. In other words, the Independence model assumes all relationships between the observed variables are zero.

5.2 Tests of Absolute Fit

The chi-square test of overall model fit is labeled *Discrepancy* in this output. Its value is 76.102 with 6 degrees of freedom, returning a probability value of less than .001 that a chi-square value this large or larger would be obtained by chance if the null hypothesis that the model fits the data is true.

The 6 degrees of freedom represent the level of overidentification of the model. Employing the earlier formula, there are 6 observed variables, so there are [6(6+1)]/2 = 21 available degrees of freedom. There are 6 residual variance estimates, 3 factor variances, 3 path coefficients, and 3 factor loadings – there are 15 parameters estimated. 21 degrees of freedom – 15 estimated parameters = 6 available degrees of freedom for the chi-square test of overall model fit.

Since the probability value of the chi-square test is smaller than the .05 level used by convention, you would reject the null hypothesis that the model fits the data. This conclusion is not good news for the researcher who hopes to fit this model to the dataset used in the example.

5.3 Tests of Relative Fit

Because the chi-square test of absolute model fit is sensitive to sample size and non-normality in the underlying distribution of the input variables, investigators often turn to various descriptive fit statistics to assess the overall fit a model to the data. In this framework, a model may be rejected on an absolute basis, yet a researcher may still claim that a given model outperforms some other baseline model by a substantial amount. Put another way, the argument researchers make in this context is that their chosen model is substantially less false than a baseline model, typically the independence model. A model that is parsimonious, and yet performs well in comparison to other models may be of substantive interest.

For example, the Tucker-Lewis Index (*TLI*) and the Comparative Fit Index (*CFI*) compare the absolute fit of your specified model to the absolute fit of the Independence model. The greater the discrepancy between the overall fit of the two models, the larger the values of these descriptive statistics.

A separate block of the output displays parsimony adjusted fit statistics. These fit statistics are similar to the adjusted R^2 in multiple regression analysis: the parsimony fit statistics penalize large models with many estimated parameters and few leftover degrees of freedom.

The fit output contains a large array of model fit statistics. All are designed to test or describe overall model fit. Each researcher has his or her favorite collection of fit statistics to report.

Commonly reported fit statistics are the chi-square (labeled *Discrepancy* in the output shown above), its degrees of freedom (*DF*), its probability value (*P*), the Tucker-Lewis Index (*TLI*), and the Root Mean Square Error of Approximation (*RMSEA*) and its lower and upper confidence interval boundaries. There is also a Standardized Root Mean Residual (*Standardized RMR*) available through the **Tools, Macro** menu, but it is important to note that this fit index is only available for complete datasets (it will not be printed for databases containing incomplete data).

Various rules of thumb for each of these fit statistics exist. These rules of thumb change as statisticians publish new simulation studies that further document the behavior of various measures of fit. The chi-square test is an *absolute* test of model fit: If the probability value (*P*) is below .05, the model is rejected. The other measures of fit are descriptive. Hu and Bentler (1999) recommend RMSEA values below .06 and Tucker-Lewis Index values of .95 or higher. Since the RMSEA for this model is .11 and the Tucker-Lewis Index value is .92, the model does not fit well according to the descriptive measures of fit.

The final column in the *Fit Measures* table is labeled *Macro* and contains the name of the corresponding macro variable for each fit statistic reported by AMOS in the *Fit Measures* table. These macro variables may be specified as part of the path diagram display if you want to display a fit index value as part of the AMOS diagram output. For example, you could use the *CMIN* macro variable to display the fitted model's chi-square on the diagram. Macro variables are used in the diagram's title, which you can specify by selecting the **Title** button from the AMOS toolbar.

Formulas for the discrepancy fit functions used to generate the chi-square test of overall model fit and the descriptive model fit statistics may be found in the AMOS manual and also in the AMOS program help files.

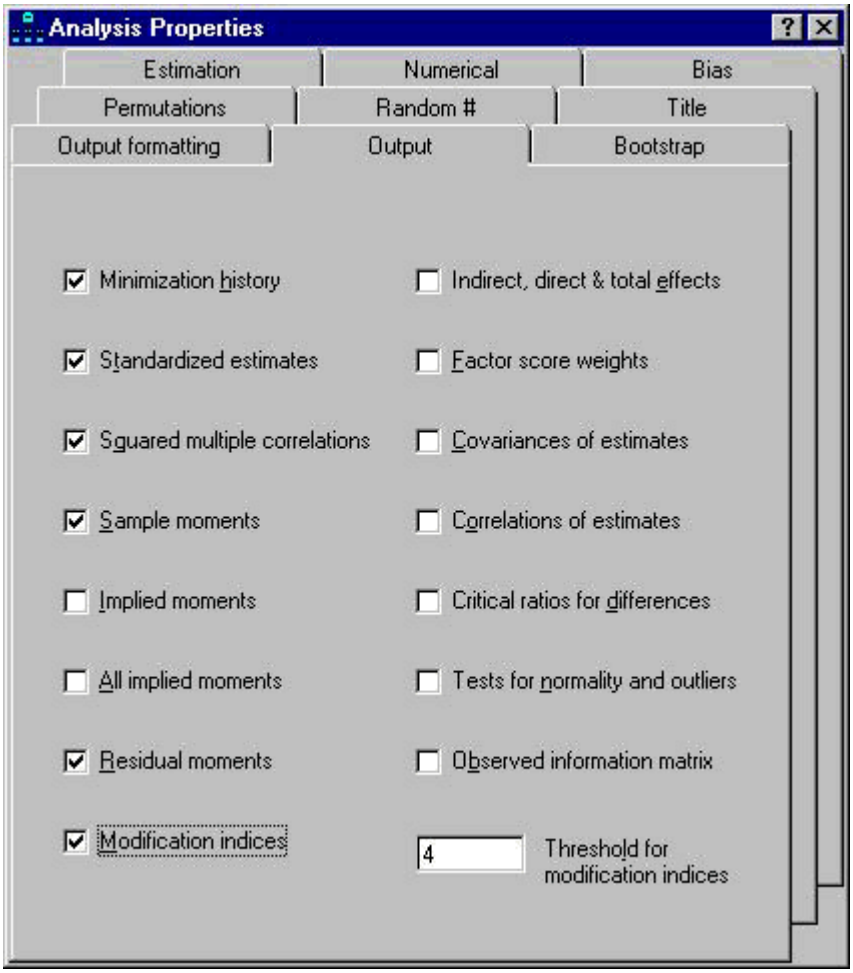
5.4 Modifying the Model to Obtain Superior Goodness of Fit

It is rare that a model fits well at first. Sometimes model modification is required to obtain a better-fitting model. AMOS allows for the use of *modification indices* to generate the expected reduction in the overall model fit chi-square for each possible path that can be added to the model. To request modification index output, select the **Modification Indices** check box in the **Output tab** in the **Analysis Properties** window.

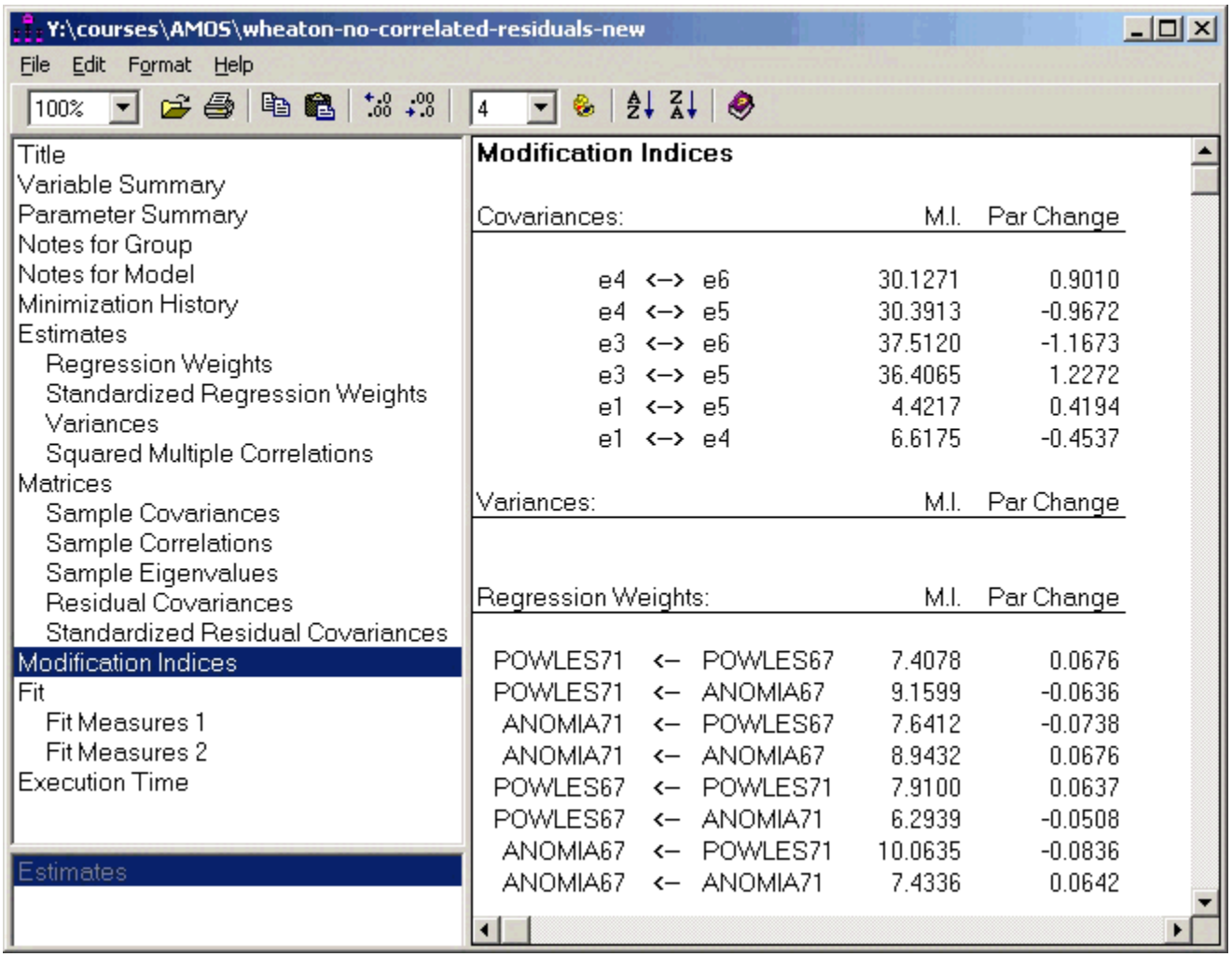
View/Set

Analysis Properties

Output



The **Threshold for Modification Indices** allows you to specify what level of chi-square change is required for a path to be included in the modification index output. The default value is 4.00 because it slightly exceeds the tabled critical value of a chi-square distribution with one degree of freedom: 3.84. Any additional parameter estimated by AMOS should result in an expected reduction in the model chi-square of at least 3.84. The modification index results appear below.



All possible variances were estimated, so there are no unmodeled variances that could be estimated in a modified or revised model. Thus, the *Variances* section contains no model modification information. There are, however, possible regression weights and covariances that can be incorporated into a re-specified model that would result in substantial changes in the model fit chi-square test statistic.

The largest modification index values are found in the first four pairs of residual covariances. For example, the covariance of *e4* with *e6* is expected to be .901 if you were to re-specify the model with that covariance added and then refit the model. That model's chi-square test of overall fit should be approximately 30.127 units lower than the present model's value of 76.102.

Should you allow the four pairs of error covariances to be estimated in a modified model? From looking at modification index results, the answer appears to be "Yes", but it is wise to reconsider the conceptual implications of model modification before you proceed further. It is important to understand that when you modify a model based upon the modification index output, you are:

1. Re-specifying your model

2. Re-specifying your model based on sample-dependent results

Any time you re-specify or modify your model, you are implicitly changing its meaning in some fundamental way. In many instances, a change in model specification results in a trivial or unimportant corresponding alteration of the model's substantive meaning, but in other cases model modification can foreshadow a strong shift in the model's meaning from a theoretical standpoint. Therefore, it is crucially important to think through each proposed model modification and ask yourself if making the modification is theoretically consistent with your research goals.

A second consideration to take into account when you modify a model is that you are relying on the empirical data rather than theory to help you specify the model. The more empirically-based modifications you incorporate into your final model, the less likely the model is to replicate in new samples of data. For these reasons, you should modify your models based upon theory as well as the empirical results provided by the modification indices.

As a practical consideration, it is also worth noting that AMOS provides modification index output only when complete data are input into the program. In other words, you cannot obtain modification index information when you use missing data with AMOS.

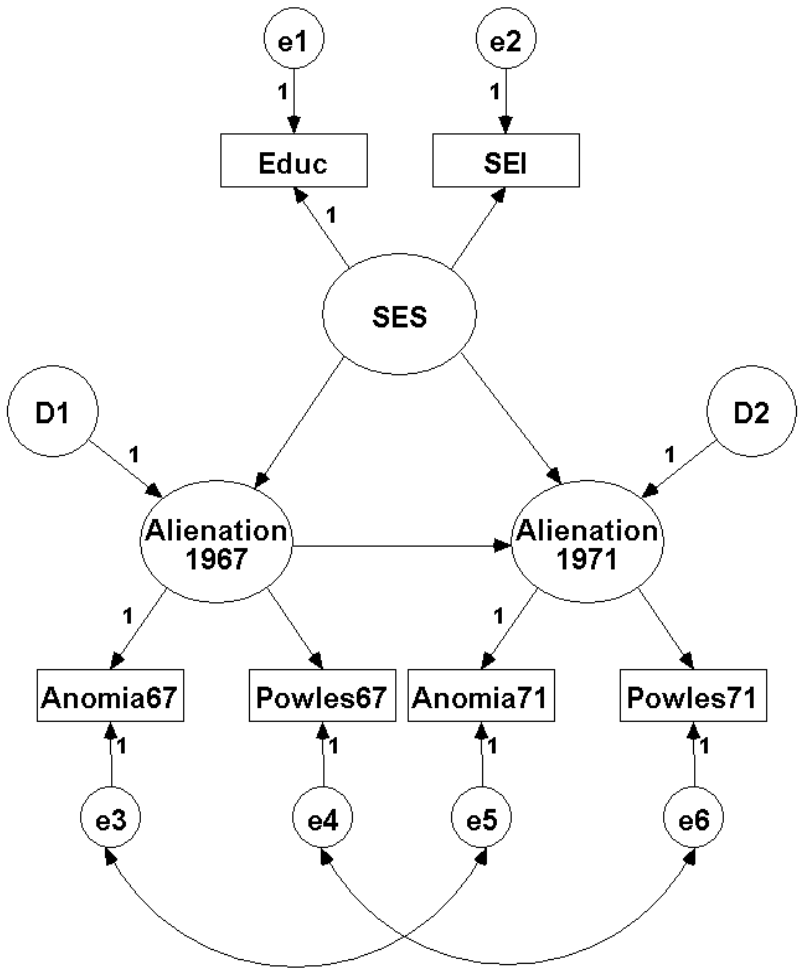
Can you apply these principles to the model from the current example? Yes. In examining the model, you can see that there are two instruments that have multiple measurements: Anomia and Powerlessness are measured in 1967 and 1971. Since these data come from the same research participants, it seems reasonable to conclude that there may be shared variance between Anomia in 1967 and Anomia in 1971 that is not captured by the present model. Similarly, there may also be shared variance between Powerlessness in 1967 and Powerlessness in 1971 that is not accounted for in the present model.

You can correlate the residuals of these two sets of variables to incorporate these sets of shared variance into the model. To correlate the residuals, return to the AMOS Graphics window and select the **Draw Covariances** tool, represented by a double-headed arrow.

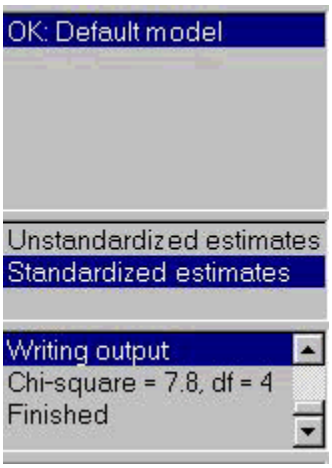
Draw a correlation between *e3*, the error term for Anomia in 1967, and *e5*, the error term for Anomia in 1971. Repeat the process for the Powerlessness residuals.

Hint: You can alter the amount of curvature in the correlation lines by using the Shape Change tool.

The modified model appears below.

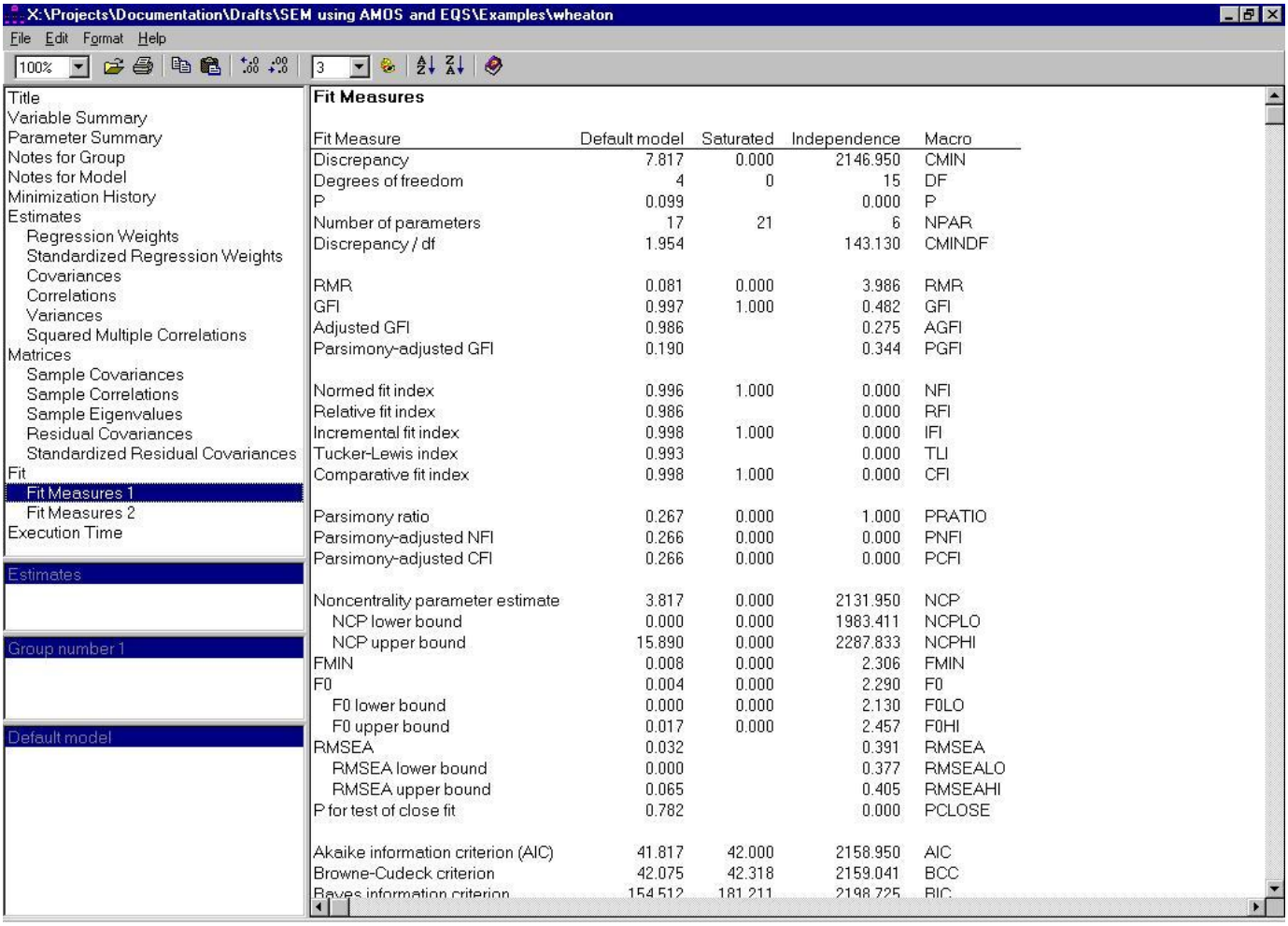


AMOS displays this result when this model is run:



Inclusion of the two correlated residuals results in a substantial drop in the model fit chi-square from 76.1 to 7.8. Notice that degrees of freedom of the chi-square test is reduced from six degrees of freedom to four degrees of freedom. The two degrees of freedom change occurs because each parameter estimated by the model consumes one degree of freedom; since you included two new parameters in the modified model (the two residual correlations), the new model has two fewer degrees of freedom remaining.

The model fit table produced by AMOS appears below:



Fit Measures

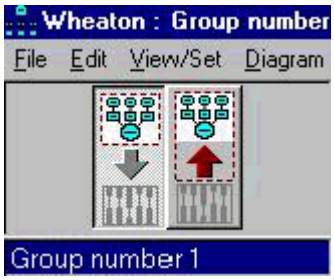
Fit Measure	Default model	Saturated	Independence	Macro
Discrepancy	7.817	0.000	2146.950	CMIN
Degrees of freedom	4	0	15	DF
P	0.099		0.000	P
Number of parameters	17	21	6	NPAR
Discrepancy / df	1.954		143.130	CMINDF
RMR	0.081	0.000	3.986	RMR
GFI	0.997	1.000	0.482	GFI
Adjusted GFI	0.986		0.275	AGFI
Parsimony-adjusted GFI	0.190		0.344	PGFI
Normed fit index	0.996	1.000	0.000	NFI
Relative fit index	0.986		0.000	RFI
Incremental fit index	0.998	1.000	0.000	IFI
Tucker-Lewis index	0.993		0.000	TLI
Comparative fit index	0.998	1.000	0.000	CFI
Parsimony ratio	0.267	0.000	1.000	PRATIO
Parsimony-adjusted NFI	0.266	0.000	0.000	PNFI
Parsimony-adjusted CFI	0.266	0.000	0.000	PCFI
Noncentrality parameter estimate	3.817	0.000	2131.950	NCP
NCP lower bound	0.000	0.000	1983.411	NCPLO
NCP upper bound	15.890	0.000	2287.833	NCPHI
FMIN	0.008	0.000	2.306	FMIN
F0	0.004	0.000	2.290	F0
F0 lower bound	0.000	0.000	2.130	F0LO
F0 upper bound	0.017	0.000	2.457	F0HI
RMSEA	0.032		0.391	RMSEA
RMSEA lower bound	0.000		0.377	RMSEALO
RMSEA upper bound	0.065		0.405	RMSEAHI
P for test of close fit	0.782		0.000	PCLOSE
Akaike information criterion (AIC)	41.817	42.000	2158.950	AIC
Browne-Cudeck criterion	42.075	42.318	2159.041	BCC
Bayes information criterion	154.512	181.211	2198.725	BIC

The chi-square value of 7.817 with four degrees of freedom is non-significant at the .05 level: its *p*-value is .099. This finding suggests that model fits the data acceptably in the population from which Wheaton et al. drew their sample. Corroborating evidence is provided by the RMSEA fit statistic – the obtained value of .032 is well below the desired .06 cutoff. Similarly, the Tucker-Lewis Index result of .993 is considerably above the .95 threshold denoting satisfactory model fit.

Once you obtain a model that fits well and is theoretically consistent, you may interpret the parameter estimates and individual tests of significance of each parameter estimate. AMOS provides two ways for you to examine parameter estimates. One method uses the path diagram output to visually display the parameter estimates while the other approach uses tables similar to those containing the overall model fit statistics.

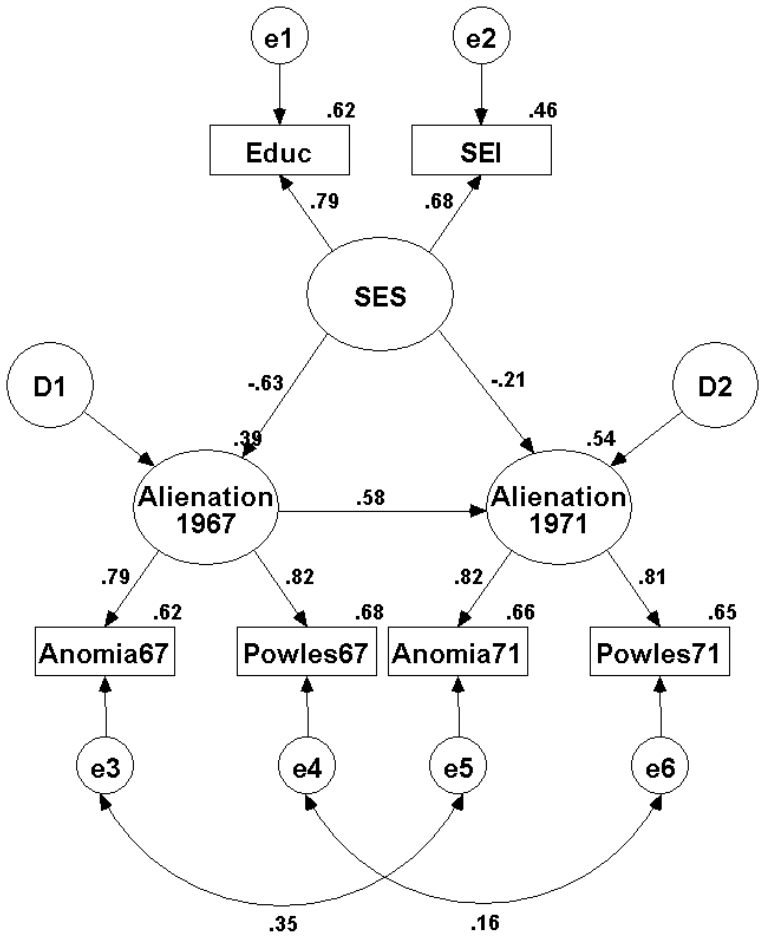
5.5 Viewing Path Diagram Output

At the top of this section of the AMOS Graphics window is an up-arrow located next to a down-arrow.



Clicking on the up-arrow will cause AMOS to display the parameter estimates. Unstandardized or standardized estimates can be chosen by clicking on the appropriate selection. Standardized estimates are selected in the figure shown below.

Clicking on the down-arrow returns you to the AMOS Graphics drawing interface, where you can modify your existing model and then re-run it, or you can open a new model or pre-existing model file. When you click on the up-arrow, the following parameter estimates are displayed as part of the output.

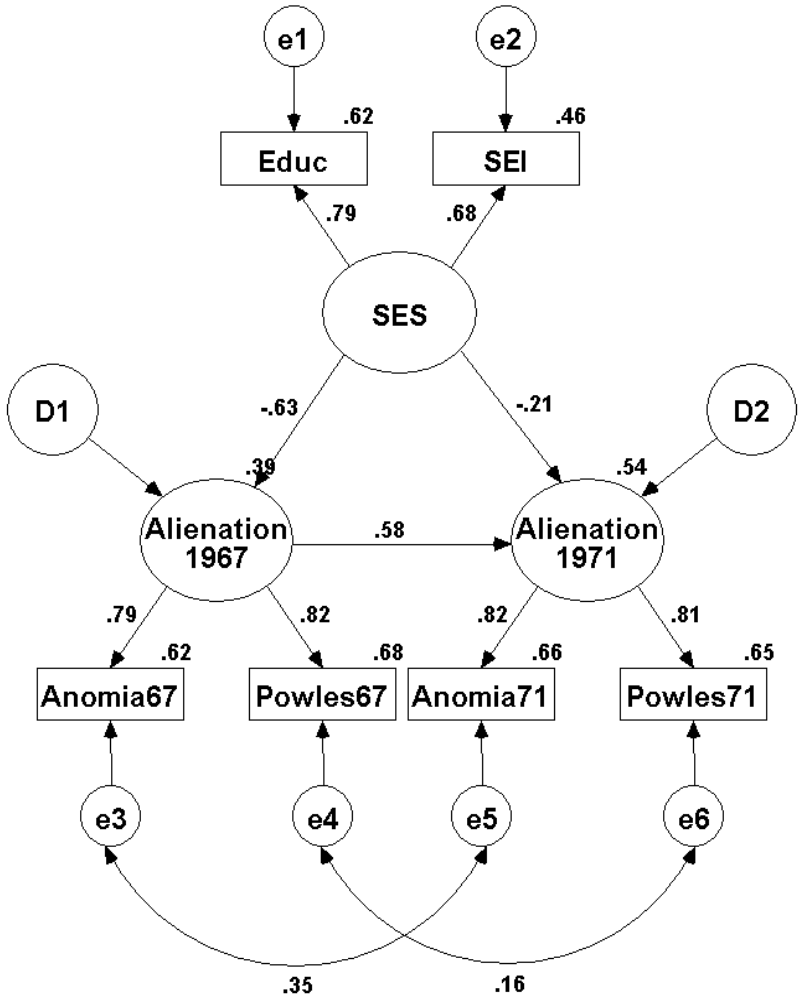


The values associated with each path are standardized regression coefficients. These values represent the amount of change in Y given a standard deviation unit change in X. (The corresponding unstandardized coefficients represent the amount of change in Y given a single raw score unit change in X).

AMOS also prints the R^2 values for each dependent or mediating variable above the variable. For example, the R^2 value for Anomia67 is .62. Although AMOS does a good job of laying out the coefficients in its default display, it may be necessary from time to time to move a particular parameter estimate value so that the drawing appears less cluttered and more easily interpretable. To move a parameter on the output diagram, use the **Move Parameter** tool.

Select the tool and move your mouse pointer over the offending variable until it is highlighted in red. Then click and pull the mouse in a direction you think would allow the parameter estimate value to be displayed more appropriately. A good choice in the present diagram is the .39 R^2

value for the *Alienation 1967* latent variable. Currently, it is partially hidden by the *SES* to *Alienation 1967* path. By dragging the parameter estimate object slightly to the right of the path object, you can see the .39 value unobstructed.



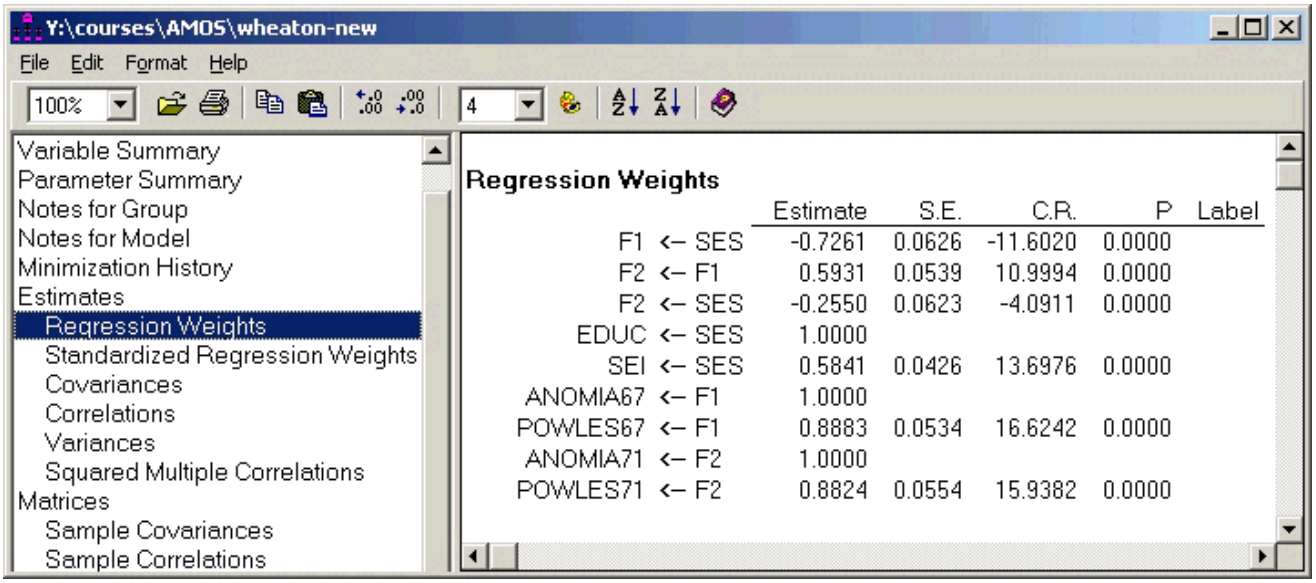
A nice feature of AMOS is its high-quality graphical output. You can take this output and copy it to the Windows clipboard. From there you can insert it into a word processor such as MS Word or a presentation package like MS PowerPoint.

This model has several interesting features worth noting. First, it contains both latent (unobserved) and manifest (observed) variables. Second, it contains both causal relationships among latent variables, represented by single-headed arrows, and correlational or bi-directional relationships among several of the residuals. These are represented by the dual-headed arrows connecting e3 with e5 and e4 with e6, respectively. As discussed above, because the two anomia and powerlessness measures are identical and measured on the same research participants across time, it makes sense that they share variance due to causes not accounted for by the alienation

latent factors. The correlations between the residuals accounts for that additional shared variance.

5.6 Significance Tests of Individual Parameters

The AMOS output also displays the unstandardized and standardized regression coefficients. The unstandardized coefficients and associated test statistics appear below. Each unstandardized regression coefficient represents the amount of change in the dependent or mediating variable for each one unit change in the variable predicting it. For example, in the figure shown below Alienation in 1967 decreases -.726 for each 1.00 increase in SES.

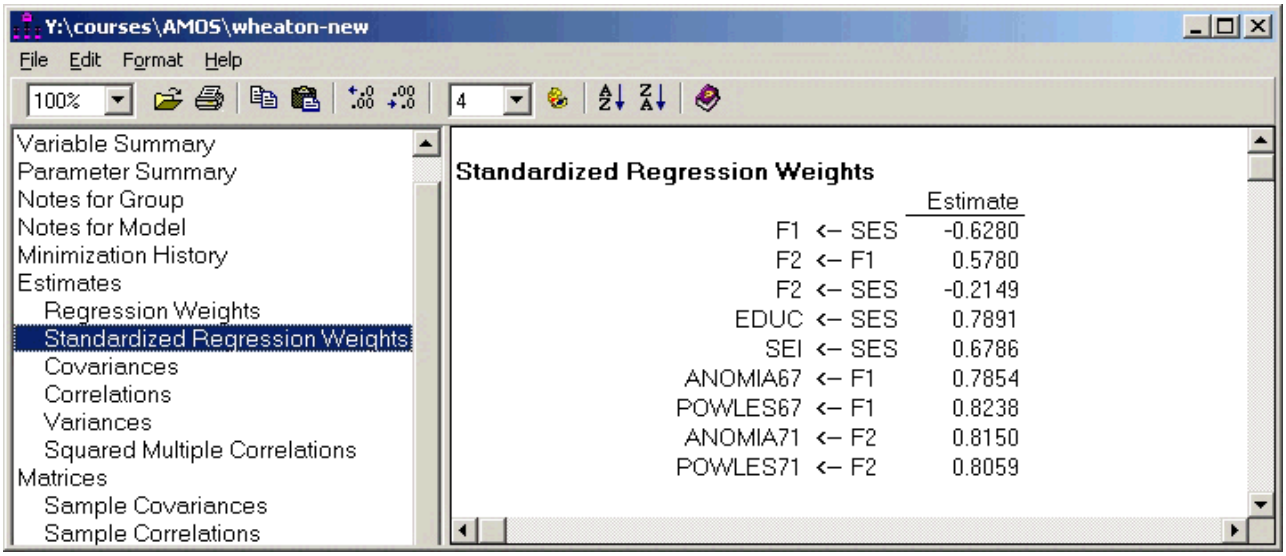


Regression Weights

		Estimate	S.E.	C.R.	P	Label
F1	<– SES	-0.7261	0.0626	-11.6020	0.0000	
F2	<– F1	0.5931	0.0539	10.9994	0.0000	
F2	<– SES	-0.2550	0.0623	-4.0911	0.0000	
EDUC	<– SES	1.0000				
SEI	<– SES	0.5841	0.0426	13.6976	0.0000	
ANOMIA67	<– F1	1.0000				
POWLES67	<– F1	0.8883	0.0534	16.6242	0.0000	
ANOMIA71	<– F2	1.0000				
POWLES71	<– F2	0.8824	0.0554	15.9382	0.0000	

The table displays the unstandardized estimate, its standard error (abbreviated **S.E.**), and the estimate divided by the standard error (abbreviated **C.R.** for Critical Ratio). The probability value associated with the null hypothesis that the test is zero is displayed under the **P** column. All of the regression coefficients in this model are significantly different from zero beyond the .01 level.

Standardized estimates allow you to evaluate the relative contributions of each predictor variable to each outcome variable. The standardized estimates for the fitted model appear below.



There is not much difference between the standardized and unstandardized coefficients in this example, probably because the units are derived from survey measurement items. By contrast, variables with very different measurement scales entered into the same model can result in sharp discrepancies between the standardized and unstandardized regression coefficient output.

Section 6: Putting it all together - A substantive interpretation of the findings

Once you have obtained a model that fits well and that is theoretically consistent and it provides statistically significant parameter estimates, you must interpret it in the light of your research questions and then distill your results in written form for publication.

Noteworthy features of this model include the negative relationship between SES and Alienation, both in 1967 and in 1971, as illustrated by the statistically significant unstandardized regression coefficients. The standardized coefficients reveal a stronger relationship between SES and Alienation in 1967 versus 1971. As one would expect, the relationship between Alienation in 1967 and Alienation in 1971 is strong as well. The measurement portion of the model is also quite good – the lowest R^2 value is .46 (*SEI*), which is a reasonable value to obtain in behavioral sciences research, and the other R^2 values are higher, indicating that the model is accounting for a large proportion of the variance in the measured items.

It is important to note that even though this model fits the data well and provides a theoretically consistent set of findings, there may be other *equivalent models* that fit the data equally well. There may also be non-equivalent alternative models that fit the data better than this model. Researchers should strive to test and rule out likely alternative models whenever possible.

Hoyle and Panter (in Hoyle, 1995) and Hatcher (1994) provide excellent discussions of how to write the results of structural equation models for publication in journals and textbooks. Hoyle and Panter provide guidelines for writing about structural equation models whereas Hatcher provides sample text from a mock write-up of the results of a SEM analysis ostensibly performed for a manuscript to be submitted for publication in a scholarly journal. Both texts' full citations appear in the References section, shown below, and are available through the UT library system.

In this course you learned

- what SEM is and where to locate SEM resources in print and on line.
- SEM nomenclature.
- the mathematical basis of SEM and its assumptions.
- how to draw a model using AMOS.
- how to run the AMOS model and evaluate several key components of the AMOS graphics and text output, including overall model fit and test statistics for individual path coefficients.
- how to modify and respecify a non-fitting model.

References

Benter, P. M. & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16(1)*, 78-117.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: John Wiley & Sons.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling.* Cary, NC: SAS Institute.

Hoyle, R. (1995). *Structural equation modeling : concepts, issues and applications.* Thousand Oaks, CA: Sage Publications.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6(1),* 1-55.

Little, R. J. A. & Rubin, D. A. (1987). *Statistical analysis with missing data.* New York NY: John Wiley & Sons.

Loehlin, J. C. (1992). *Latent variable models.* Hillsdale, NJ: Lawrence Erlbaum Publishers.

Rigdon, E. (1997). *Approaches to testing identification.* http://www.gsu.edu/~mkteer/identifi.html

Roth, P. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47,* 537-560.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences.* Mahwah, NJ: Lawrence Erlbaum Publishers.

Wheaton, B., Muthén, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in panel models. In D.R. Heise (Eds.): *Sociological Methodology.* San Fransisco: Jossey-Bass.