13
Analyzing Path Models Using SEM Programs

To this point I have used multiple regression for the analysis of path models (as well as multiple regression models). It is also possible to use dedicated structural equation modeling (SEM) programs for such analysis. We make that switch in this chapter. As you will see, the results of simple path analyses are identical using SEM or MR analysis, but SEM programs can analyze more complex models and have real advantages when analyzing overidentified models.

SEM PROGRAMS

Numerous SEM programs are available, all of which are capable of analyzing everything from simple path models through latent variable structural equation models. LISREL (Linear Structural Relations; Jöreskog & Sörbom, 1996; Mels, 2006) was the first such program and is still widely used. For additional information, go to www.ssicentral.com. Other common programs include EQS (Bentler, 1995; www.mvsoft.com) and Mplus (Muthén & Muthén,

1998–2012; www.statmodel.com). Each such program has its own advantages; Mplus, for example, has sophisticated routines for analyzing categorical variables and is perhaps the most flexible such program. They generally cost $500 to $600 for those in academia, and many of them have try-out or trial versions and reduced pricing for students. For users of R (a free statistical programing language), there are at least two free SEM add-ons for that I know of, OpenMx (http://openmx.psyc.virginia.edu/; Boker et al., 2012) and lavaan (http://lavaan.ugent.be/; Beaujean, 2014; Rosseel, 2012).

Amos and Mplus

My favorite teaching program is one called Amos (Analysis of Moment Structures; Arbuckle, 2013; www.spss.com/amos), although I also use Mplus on a regular basis. Amos uses a graphic approach and is probably the most intuitive and easiest SEM program to use. It produces attractive path diagrams (all the path models you have seen so far were produced by Amos) and can be used both to draw a path diagram and analyze it. As of this writing, student pricing for Amos is around $50 per year as an SPSS for Windows add-on (there is no Mac version). The user's guide for the most recent version is also available as a pdf document on the spss website (under product support) and you can download the program as a free try-out for 14 days. Of course, you can analyze these problems using any SEM program, so if you have another program available you may want to use it. As noted, there are also student or demo versions available of many of the commercial SEM programs.

There are numerous examples of Amos and Mplus input and output—at least one per chapter—on the website (www.tzkeith.com). Statistical programs are revised constantly, so check the website also for more up-to-date information than is contained here in the text. Whatever program you use, you should download or purchase the user's manual, which provides the basics for the use of the program. There are numerous other sources of information about various SEM programs, as well. If you use Amos, for example, I recommend you download a tutorial from http://ssc.utexas.edu/training/software-tutorials . This site also has tutorials for other SEM and general statistics programs, and there are, of course, a growing number of such resources on the web. Although I will generally use Amos to estimate subsequent models, the information presented applies to SEM programs in general.

Basics of SEM Programs

Everything you have learned about path analysis so far will transfer to Amos and other SEM programs. Figure 13.1 shows a basic SEM (Amos) version of the parent involvement model first presented in Chapter 12. As in all previous examples, rectangles represent measured variables, and ovals represent unmeasured or latent variables (in this example, the disturbances). Straight arrows represent paths, or presumed influences, and curved, double-headed arrows represent correlations (or, with unstandardized coefficients, covariances). The one new aspect of Figure 13.1 is the value of 1 beside the paths from the disturbances to the endogenous variables. These paths simply set the scale of measurement for the disturbances. Unmeasured variables have no natural scale. When we set the path from the disturbances, which are unmeasured variables, to the measured variables to 1.0, we are merely telling the SEM program that the disturbance should have the same scale as the measured variable. (In reality, any number could be used: .80, 2.0, but 1.0 is the most common and most straightforward.) We will use the same rule of thumb when we begin using other latent variables: we will set one path from each latent variable to 1.0 to set the scale of the latent variable. At a practical level, the model would be underidentified without this constraint (or some other way of setting the scale of the disturbances). Depending on which program you use, these

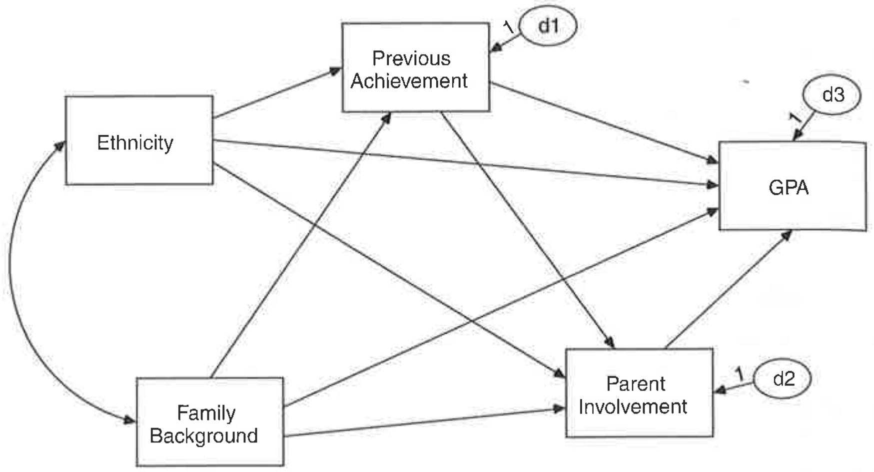

Figure 13.1 Parent Involvement model from Chapter 12, as drawn in the Amos SEM program.

paths from disturbances to endogenous variables may be set to 1 automatically and invisibly (for example, this happens by default in MPlus).

We could also set the scale by fixing the variance of the disturbance to 1.0; all substantive results would be the same. In fact, this is exactly what we did with multiple regression, even though we did not realize that we were doing so. When we use multiple regression to estimate the paths, the variances of the disturbances are set to 1.0, and the program estimates the paths from the disturbances to the endogenous variables (when we set the path to 1.0, the program estimates the variance of the disturbance). We can choose either method with Amos; I here set the paths to 1.0 because that is the most common method.

REANALYSIS OF THE PARENT INVOLVEMENT PATH MODEL

The model shown in Figure 13.1 provides the basic input for analysis by Amos (the model is on the Web site as "PI Example 1.amw"); add data and you can conduct the analysis. Most SEM programs, including Amos, can use the correlation matrix and standard deviations as input for the analysis. The matrix for this example is saved as both an SPSS (PI matrix, listwise.sav) and an Excel file (PI matrix, listwise.xls). The matrix is also shown in Table 13.1; the

Table 13.1 Means, Standard Deviations, Sample Sizes, and Correlations among the Variables for the Parent Involvement Path Example

Variable	*Ethnic*	*Byses*	*Bytests*	*Par_inv*	*Ffugrad*
ETHNIC	1.000	.333	.330	.075	.131
BYSES	.333	1.000	.461	.432	.299
BYTESTS	.330	.461	1.000	.445	.499
PAR_INV	.075	.432	.445	1.000	.364
FFUGRAD	.131	.299	.499	.364	1.000
M	.793	.047	52.323	.059	5.760
SD	.406	.766	8.584	.794	1.450
N	811.000	811.000	811.000	811.000	811.000

variable names are as in the NELS raw data. The SPSS commands I used to create the matrix using the NELS data are in the file "create corr matrix in spss.sps."

Estimating the Parent Involvement Model via Amos

With the model and the data, we can estimate the model via Amos (or any other SEM program). The standardized output for this model is shown in Figure 13.2. Compare the results with your results from Chapter 12; with the exception of the lack of a number associated with the paths from disturbances to endogenous variables, the results should be identical. Figure 13.3 shows the unstandardized output for the model. Recall that we set the paths from disturbances to endogenous variables to 1.0 and estimated the variances of the disturbances. The numbers next to the disturbances are the estimates of their variances. The numbers above the two exogenous variables are their variances. Again, the results should match those from your regression analysis in Chapter 12.

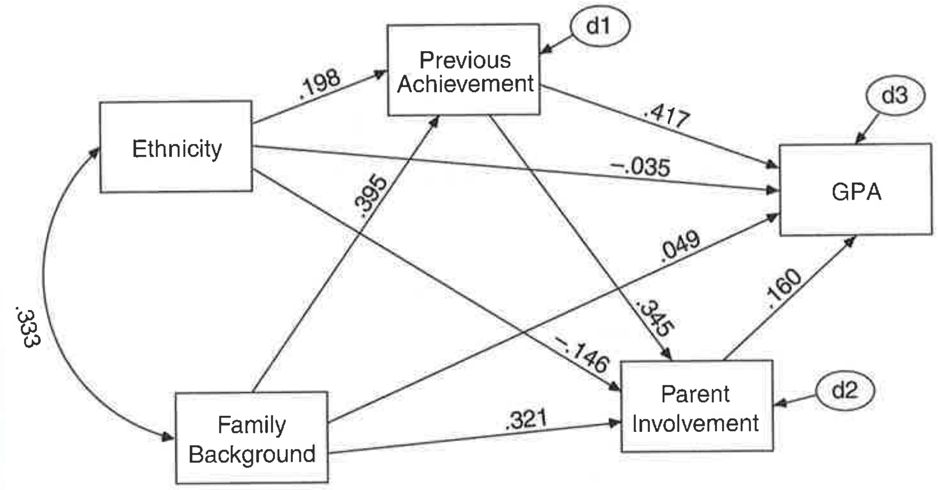

Figure 13.2 Parent Involvement model estimated via Amos. The standardized results are the same as those in Chapter 12 when the model was estimated via multiple regression.

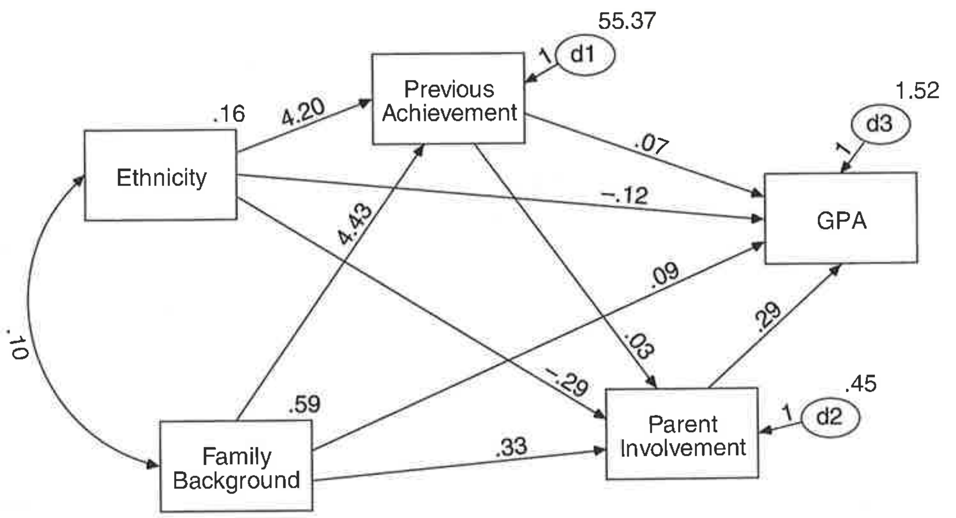

Figure 13.3 Unstandardized estimates for the Parent Involvement model.

Of course, you will get more detailed output than just these diagrams from your SEM program. Figure 13.4 shows one portion of the printout; this and all subsequent printouts show Amos output, but you will get something similar with any of the SEM programs. The top portion of the output (Regression Weights) shows the unstandardized path coefficients, listed under the column Estimate. For example, the first row shows that the unstandardized path from BYSES (Family Background) to bytests (Previous Achievement) is 4.431 (it is possible to have Amos list the longer variable labels in addition to the variable names, but just the names are shown in this output). The S.E. column shows the standard errors of the

Regression Weights

	Estimate	S.E.	C.R.	P	Label
bytests <--- BYSES	4.431	.362	12.229	***	
bytests <--- Ethnic	4.195	.684	6.131	***	
par_inv <--- bytests	.032	.003	10.034	***	
par_inv <--- Ethnic	-.286	.063	-4.525	***	
par_inv <--- BYSES	.333	.036	9.345	***	
ffugrad <--- bytests	.070	.006	11.406	***	
ffugrad <--- par_inv	.292	.064	4.528	***	
ffugrad <--- BYSES	.093	.069	1.354	.176	
ffugrad <--- Ethnic	-.124	.117	-1.057	.290	

Standardized Regression Weights

	Estimate
bytests <--- BYSES	.395
bytests <--- Ethnic	.198
par_inv <--- bytests	.345
par_inv <--- Ethnic	-.146
par_inv <--- BYSES	.321
ffugrad <--- bytests	.417
ffugrad <--- par_inv	.160
ffugrad <--- BYSES	.049
ffugrad <--- Ethnic	-.035

Covariances

	Estimate	S.E.	C.R.	P	Label
Ethnic <--> BYSES	.103	.011	8.996	***	

Correlations

	Estimate
Ethnic <--> BYSES	.333

Variances

	Estimate	S.E.	C.R.	P	Label
Ethnic	.164	.008	20.125	***	
BYSES	.586	.029	20.125	***	
d1	55.370	2.751	20.125	***	
d2	.452	.022	20.125	***	
d3	1.521	.076	20.125	***	

Figure 13.4 Output from the SEM program (Amos) showing unstandardized coefficients (regression weights), their standard errors, and critical ratios, along with standardized coefficients.

coefficients, the column labeled C.R. (for critical ratio) shows the *t*'s for each coefficient. (Recall that $t = \text{coefficient}/SE_{\text{coefficient}}$ and that with large samples *t*'s greater than approximately 2 are statistically significant. The values are actually *z* statistics, but they are essentially the same with the sample sizes we are using.) The column labeled P shows the probability associated with each path, with values less than .001 indicated by ***. The next portion of the figure (Standardized Regression Weights) shows the standardized paths. Again, the output should match the SPSS output from Chapter 12. This portion is followed by the covariance and correlation between the two exogenous variables, the variances of the two exogenous variables and the variances of the disturbances of the three endogenous variables.

SEM programs will also produce tables of direct, indirect, and total effects for both the standardized and unstandardized solution. The tables for the current example are shown in Figure 13.5. The tables are read from column to row; thus the total unstandardized effect of

Total Effects

	BYSES	Ethnic	bytests	par_inv
bytests	4.431	4.195	.000	.000
par_inv	.474	-.153	.032	.000
ffugrad	.544	.127	.080	.292

Standardized Total Effects

	BYSES	Ethnic	bytests	par_inv
bytests	.395	.198	.000	.000
par_inv	.458	-.078	.345	.000
ffugrad	.287	.035	.472	.160

Direct Effects

	BYSES	Ethnic	bytests	par_inv
bytests	4.431	4.195	.000	.000
par_inv	.333	-.286	.032	.000
ffugrad	.093	-.124	.070	.292

Standardized Direct Effects

	BYSES	Ethnic	bytests	par_inv
bytests	.395	.198	.000	.000
par_inv	.321	-.146	.345	.000
ffugrad	.049	-.035	.417	.160

Indirect Effects

	BYSES	Ethnic	bytests	par_inv
bytests	.000	.000	.000	.000
par_inv	.141	.134	.000	.000
ffugrad	.450	.251	.009	.000

Standardized Indirect Effects

	BYSES	Ethnic	bytests	par_inv
bytests	.000	.000	.000	.000
par_inv	.136	.068	.000	.000
ffugrad	.238	.070	.055	.000

Figure 13.5 Total, indirect, and direct effects of variables on each other in the Parent Involvement model.

Family Background (BYSES) on GPA (ffugrad), as shown in the bottom left of the first table, is .544. Take some time to compare these results with those from the previous chapter.

It is also possible to evaluate the statistical significance of the indirect and total effects; in Amos this is done through a bootstrapping procedure. (*Bootstrapping* is a procedure in which one takes repeated, smaller random samples of an existing sample. With bootstrapping, it is possible to develop empirical estimates of standard errors of any parameter, even, for example, standard errors of standard errors. Recall that in previous discussions of mediation, aka indirect effects, I have said that there are better ways of assessing the statistical significance of indirect effects than the Sobel test. Bootstrapping is such a method.) Figure 13.6, for example, shows the indirect effects for the variables in the Parent Involvement model, followed by their standard errors. You can use this information to calculate the *t* values for

Indirect Effects

	BYSES	Ethnic	bytests	par_inv
bytests	.000	.000	.000	.000
par_inv	.141	.134	.000	.000
ffugrad	.450	.251	.009	.000

Standardized Indirect Effects

	BYSES	Ethnic	bytests	par_inv
bytests	.000	.000	.000	.000
par_inv	.136	.068	.000	.000
ffugrad	.238	.070	.055	.000

Indirect Effects - Standard Errors

	BYSES	Ethnic	bytests	par_inv
bytests	.000	.000	.000	.000
par_inv	.018	.025	.000	.000
ffugrad	.045	.065	.002	.000

Standardized Indirect Effects - Standard Errors

	BYSES	Ethnic	bytests	par_inv
bytests	.000	.000	.000	.000
par_inv	.017	.013	.000	.000
ffugrad	.023	.018	.013	.000

Indirect Effects - Two Tailed Significance (BC)

	BYSES	Ethnic	bytests	par_inv
bytests	...	...	...	...
par_inv	.001	.001	...	...
ffugrad	.002	.002	.002	...

Standardized Indirect Effects - Two Tailed Significance (BC)

	BYSES	Ethnic	bytests	par_inv
bytests	...	...	...	...
par_inv	.002	.001	...	...
ffugrad	.002	.002	.002	...

Figure 13.6 Indirect effects (both unstandardized and standardized) for the Parent Involvement model, their standard errors, and statistical significance.

each indirect effect to determine its statistical significance; this is done in the bottom of the figure. Thus, SEM programs allow a more direct test of the statistical significance of mediation than do most regression results (see in Chapter 8 the section on Mediation). Amos also provides the *standardized indirect and total* effects and their standard errors and statistical significance (the standardized indirect effects are shown in the figure).

ADVANTAGES OF SEM PROGRAMS

Overidentified Models

Figure 13.7 shows a potential model of the effect of Homework on GPA. The data are from NELS (the larger NELS data, not those on the Web site). For this model, Ethnicity, Family Background, and Previous Achievement were measured in eighth grade and are defined in the way we have in the past (Ethnicity = majority vs. minority, Family Background = BYSES, Previous Achievement = BYTests). Homework was based on student reports of time spent on homework in each academic area, measured in both eighth and tenth grades; it may be considered a measure of average homework over time. Grades are students' GPAs (English, Math, Science, and Social Studies) from 10th grade.

Note that several potential paths are not drawn: there are no paths from Ethnicity and Family Background to Grades. Just as it means something to draw a path, it means something to not draw a path and, in fact, it is often a *stronger statement* than drawing a path. When we draw a path, we are stating that one variable may have some effect on another. What the *lack of path* from Family Background to Grades means is that I believe the path from Background to Grades is a value of zero. Indeed, not drawing a path is the same as drawing a path and fixing or constraining that path to a value of zero. This model also makes explicit the notion that the only way Ethnicity and Family Background affect Grades is through Homework and Previous Achievement, that they have no direct effect on Grades, only indirect effects through other variables in the model. I developed this hypothesis in the usual way, based on previous research and logic. Indeed, you will even find support for the exclusion of paths

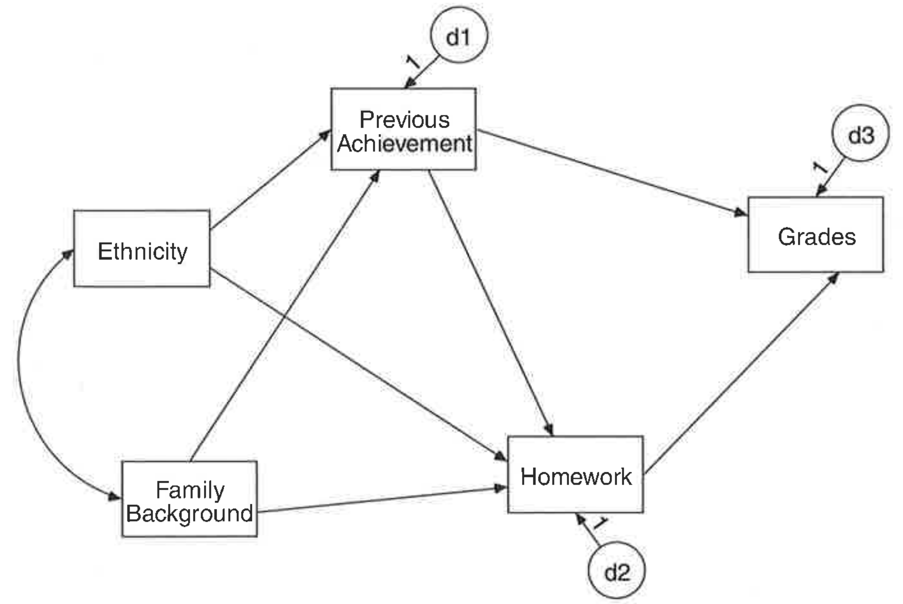

Figure 13.7 Overidentified model testing the effects of Homework on students' Grades in High School.

from Ethnicity and Family Background to Grades based on our Parent Involvement models, which showed only small direct effects for these variables on Grades.

You know from Chapter 11 that this is an overidentified model, meaning that we have more information than we need to solve for the paths. Note that there are 10 correlations among variables, but we are solving for only eight parameters (seven paths and one correlation). Recall also that if we were solving for the paths using algebra we could come up with multiple formulas for solving some of the paths. I argued in Chapter 11 that this approach may have advantages, because similarity in path estimates calculated two different ways can give us additional confidence in our model, whereas dissimilarity might make us wonder about the veracity of the model.

One advantage of SEM programs is that they provide this type of feedback about overidentified models. The method is not as described above; the programs do not estimate the paths several different ways and allow you to compare the different estimates. Instead, the programs compare matrices and provide measures of the fit of the model to the data. We'll see how this process works when we analyze the model in Figure 13.7.

The data (correlation matrix, standard deviations, means, and *N*) are contained in both an Excel and an SPSS file ("homework overid 1.xls" and "homework overid 1.sav"); the data are also shown in Table 13.2.[1] The model shown in the figure was used as input to Amos and is in the file "homework path 1.amw" on the accompanying Web site.

Figure 13.8 shows the solved, standardized path model. Using our rules of thumb, it appears Homework has a moderate effect (.15) on 10th-grade GPA. Previous Achievement had a strong effect on Homework, suggesting a "rich get richer" sort of effect: students who achieve at a high level do more homework, and this homework, in turn, improves their subsequent school performance. Family Background also had a moderate effect on Homework, but Ethnicity had no substantive effect.

The file includes standard deviations, sample sizes, and correlations. Means are included but are not required or analyzed.

How can we use the overidentification status of the model to assess the model? Recall how we solved for the paths in our first example of path analysis: through the use of algebra, the tracing rule, and the correlations among the variables. Amos is essentially doing the same thing here: the model specification and the correlation matrix (actually the covariance matrix, but we will address this point later) were used as input, and the program used these pieces of information to solve for the paths. If we can solve for the paths using the

Table 13.2 Contents of the Excel file for the homework path example

*rowtype_*	*varname_*	*Ethnic*	*FamBack*	*PreAch*	*Homework*	*Grades*
n		1000	1000	1000	1000	1000
corr	Ethnic	1				
corr	FamBack	0.3041	1			
corr	PreAch	0.3228	0.4793	1		
corr	Homework	0.0832	0.2632	0.2884	1	
corr	Grades	0.1315	0.2751	0.489	0.2813	1
stddev		0.4186	0.8311	8.8978	0.8063	1.479
mean		0.7282	0.0025	52.0039	2.565	5.7508

The matrix is in the format required for analysis in Amos. These include the rowtype_ and varname_ columns, and the n, corr, and stddev rows (the mean row is not required at this stage of our adventures).

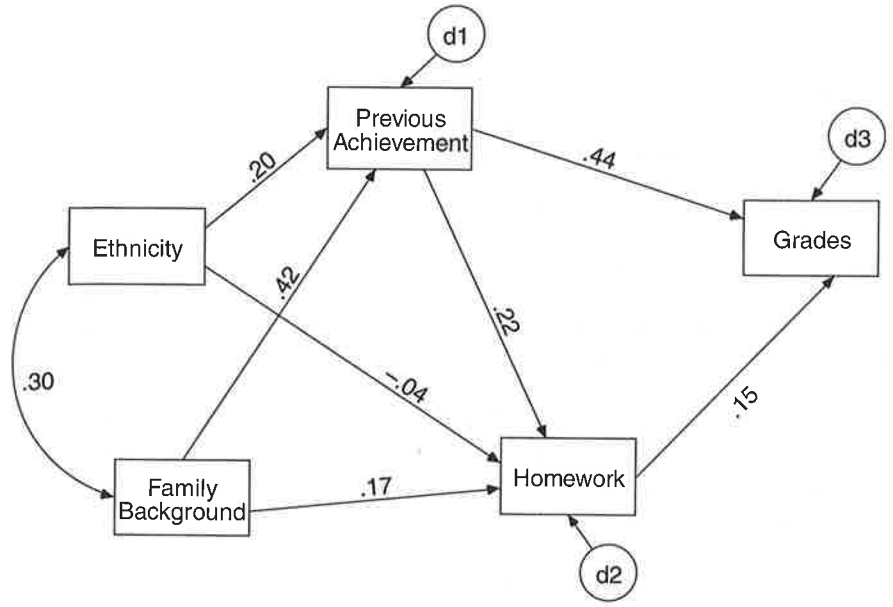

Figure 13.8 Standardized output for the Homework model.

Sample Correlations

	Ethnic	FamBack	PreAch	Homework	Grades
Ethnic	1.000				
FamBack	.304	1.000			
PreAch	.323	.479	1.000		
Homework	.083	.263	.288	1.000	
Grades	.132	.275	.489	.281	1.000

Implied (for All Variables) Correlations

	Ethnic	FamBack	PreAch	Homework	Grades
Ethnic	1.000				
FamBack	.304	1.000			
PreAch	.323	.479	1.000		
Homework	.083	.263	.288	1.000	
Grades	.156	.253	.489	.281	1.000

Figure 13.9 The sample (input) correlation matrix compared to the matrix implied by the Homework model.

correlations, why can't we do the reverse: solve for the correlations using the paths? In fact, we can do exactly that. You could, and SEM programs do, use the solved path model (e.g., Figure 13.8) to calculate an expected, or predicted, correlation matrix, the matrix implied by the model.[2] With overidentified models, this implied matrix (also known as the predicted matrix) and the input matrix will differ to some degree. The actual correlation matrix and the implied correlation matrix from the Amos output are shown in Figure 13.9. Notice that most of the correlations are the same, but that the values in the lower left—the correlations of Grades with Ethnicity and Family Background—differ slightly between the actual and the implied matrices. SEM programs use this degree of similarity or nonsimilarity between the two matrices to assess and measure the fit of the model to the data. This information is also useful for diagnosing and correcting model problems.

Correlations versus Covariances

Before going any further, it is time to augment our thinking about correlation matrices with the additional considerations of covariance matrices. Most SEM programs are set up to analyze covariance rather than correlation matrices. For some SEM problems you will get the same substantive answer no matter which type of matrix you analyze, but for others you should analyze covariance matrices (see Cudeck, 1989, or Steiger, 2001, for further discussion about this issue). An easy solution is simply to get in the habit of analyzing covariance, rather than correlation, matrices. (An alternative is to use a program, such as SEPATH, a part of the Statistica package, designed specifically to analyze correlation matrices.)

Recall from Chapter 1 that we can easily calculate covariances from correlations if we know the variances or standard deviations of the variables, because $CoV_{xy} = r_{xy} \times SD_xSD_y$. Indeed, this is what Amos did; we input the correlations and standard deviations, and the program generated the covariance matrix from that input. The covariance matrix is shown at the top portion of Figure 13.10. The covariances are shown below the diagonal, and the variances are shown in the diagonal. Another way of thinking of covariance versus correlation matrices is to recall that correlation matrices are standardized covariance matrices, with all variables converted to *z* scores.

Model Fit and Degrees of Freedom

SEM programs, then, generally compare the actual covariance matrix to the implied covariance matrix. Some of the relevant output from Amos is shown in Figure 13.10: the actual covariance matrix, the implied matrix, and the residual covariance matrix. The residual covariance matrix is the result of subtracting the implied matrix from the actual matrix; intuitively, large differences between these matrices and large residuals should signal problems with the model. More helpful are the *standardized* residuals, in which the residuals have been converted to a common, standardized metric. These are standardized like z-scores. Thus, one rule of thumb is that standardized residuals larger than 2 signal an area of misfit in the model. The values of standardized residuals are dependent on sample size, however, so large sample will produce many more large values than will small samples. Thus, a better rule of thumb is to focus on the largest values in this matrix when the more global fit statistics suggest a lack of fit.

Table 13.3 shows another useful, related matrix: the differences between the actual correlations and those implied by the model (the Sample Correlations from Figure 13.9 minus the Implied Correlations). As we will see, this matrix can also be useful for isolating and understanding problems with models. Kline (2011), for example, advises to examine further variables with such "correlation residuals" greater than .10 or less than –.10. (chap. 5). This matrix is not produced in all SEM programs (it is not produced in Amos or Mplus, for example), but it is easily generated in Excel by subtracting the values in the implied correlation matrix from those in the actual correlation matrix. As shown in the table, the correlation between Ethnicity and Grades that is implied by the Homework was slightly larger than the actual correlation; in contrast, the actual correlation between Family Background and Grades was slightly larger than the correlation implied by the Homework model.

Although these residual matrices are not particularly useful in the present example in which only two paths have been constrained, as we begin to focus on more complex and

Sample Covariances

	Ethnic	FamBack	PreAch	Homework	Grades
Ethnic	.175				
FamBack	.106	.690			
PreAch	1.201	3.541	79.092		
Homework	.028	.176	2.067	.649	
Grades	.081	.338	6.429	.335	2.185

Implied (for All Variables) Covariances

	Ethnic	FamBack	PreAch	Homework	Grades
Ethnic	.175				
FamBack	.106	.690			
PreAch	1.201	3.541	79.092		
Homework	.028	.176	2.067	.649	
Grades	.097	.311	6.429	.335	2.185

Residual Covariances

	Ethnic	FamBack	PreAch	Homework	Grades
Ethnic	.000				
FamBack	.000	.000			
PreAch	.000	.000	.000		
Homework	.000	.000	.000	.000	
Grades	–.015	.027	.000	.000	.000

Standardized Residual Covariances

	Ethnic	FamBack	PreAch	Homework	Grades
Ethnic	.000				
FamBack	.000	.000			
PreAch	.000	.000	.000		
Homework	.000	.000	.000	.000	
Grades	–.776	.662	.000	.000	.000

Figure 13.10 Sample and implied covariance matrices and residual and standardized residual matrices for the Homework model.

Table 13.3 Differences between the Actual Correlations and those Implied by the Homework Model

	Ethnic	*FamBack*	*PreAch*	*Homework*	*Grades*
Ethnic					
FamBack	0				
PreAch	0	0			
Homework	0	0	0		
Grades	–.025	.022	0	0	

latent variable models the standardized residual covariances and the residual correlations will be useful for determining *where* there is misfit in our models. I focused on the actual, implied, and residual matrices now, however, because this difference between the actual and implied covariance matrix is the source of other measures of the fit of the model.

We can and will quantify the *degree* to which a model is overidentified. The current model has two paths that could have been drawn to make the model just-identified (paths from Ethnicity and Family Background to Grades). The model thus has two degrees of freedom. More exactly, we can calculate the degrees of freedom using the following steps:

1. Calculate the number of variances and covariances in the matrix using the formula $[p \times (p + 1)]/2$ where p is equal to the number of variables in the model. For the current model, there are 15 variances and covariances: $[5 \times (5 + 1)]/2 = 15$.
2. Count the number of parameters that are estimated in the model. Don't forget covariances between exogenous variables, variances of the exogenous variables, and variances of the disturbances. For the current model, we estimated seven paths, one covariance between the exogenous variables, the variances of the two exogenous variables, and the variances of the three disturbances, for a total of 13 estimated parameters.
3. The degrees of freedom are calculated by subtracting the number of estimated parameters from the number of variances and covariances. The present model has two degrees of freedom $(15 - 13 = 2)$.

The degrees of freedom for a model provide information about the degree to which the overall model is overidentified. The degrees of freedom also provide a handy index of the parsimony of the model. In science, we value parsimony: if two explanations for a phenomenon are equally good (or, in SEM, fit equally well), we generally prefer the simplest or more parsimonious explanation. Degrees of freedom are an index of the parsimony of a path model: the more degrees of freedom, the more values constrained (to zero or some other value) prior to estimation, and thus the greater the parsimony.

The difference between the actual and implied matrices provides evidence of the degree to which the model is a good explanation of the data. This difference is used to generate a multitude of fit statistics or fit indexes for overidentified models. There are literally dozens of such fit indexes, with different indexes focusing on slightly different aspects of fit. We will focus on a few common such indexes here; there are also numerous sources for more information about fit indexes (e.g., Fan, Thompson, & Wang, 1999; Hoyle, 1995; Hu & Bentler, 1998, 1999; Marsh, Hau, & Wen, 2007; see also David Kenny's web pages for excellent and up-to-date advice on fit statistics: http://davidakenny.net/cm/fit.htm).

Chi-square (χ^2) is the most commonly reported measure of fit.[3] Chi-square has the advantage of allowing a statistical test of the fit of the model; it can be used with the degrees of freedom to determine the probability that the model is "correct" (to be explained later). Interestingly, in SEM we want a small χ^2 and one that is not statistically significant. For our current example, $\chi^2 = 2.166$, with 2 *df* and a probability of .338. What does this mean? It means that the actual and the implied matrix are not statistically significantly different from one another, and thus the model and the data are consistent with one another. If the model and the data are consistent, the model *could have* generated the data and thus may provide a good approximation of how the phenomenon being studied works. In other words, the model may approximate reality, it may be "correct." Given all the "mays" and "coulds" in this explanation, you may be disappointed; this is hardly the kind of evidence of the quality of the model you were hoping for! Sorry; fit statistics do *not* prove that a model is true and do *not* prove causality. If the fit indexes are good, they suggest that a model may provide a reasonable, tentative explanation of the data. I'll simply note that this is better than nothing and

more feedback than we've had in previous chapters about the quality of our explanations of our data.

Figure 13.11 shows the fit indexes output by Amos; other SEM programs will provide an equally intimidating listing of indexes of fit, many of which will be the same (although some may be labeled differently). Focus on the first few rows and columns. The model that is being estimated (i.e., the model in Figure 13.7) is labeled the Default model. The first column of numbers shows the number of parameters (NPAR) that are estimated in the model (remember we calculated 13 parameters being estimated), and the second shows the χ^2 (labeled CMIN, a value of 2.166). These are followed by the degrees of freedom (2) and the probability associated with the χ^2 and *df* (.338).

The rows labeled Saturated model pertain to a just-identified model. A just-identified model will estimate 15 parameters and thus have zero *df*. With a just-identified model, the implied covariance matrix will be identical to the actual matrix, and thus χ^2 associated with a just-identified model is equal to zero. In other words, a just-identified model will provide a perfect fit to the data. Why not, then, continue to estimate just-identified models, as we have done previously? The reason, again, is that we value parsimony. An overidentified model is more parsimonious than a just-identified model; our present overidentified model fits as well as a just-identified model (another interpretation of the statistically not significant χ^2). Because this model fits as well but is more parsimonious, it is a preferable model from a scientific standpoint.

The rows labeled Independence model refer to a model in which the variables in the model are assumed to be independent of one another. This model, also called a *null* model, could be represented by the five variables, with no paths or correlations drawn (and thus for this model all we would estimate would be the five variable variances). It could also be represented by constraining all paths and correlations in the current model to zero. Again, the null model assumes the variables are unrelated. The saturated and independence models essentially provide two endpoints with which we can compare our theoretical model. The saturated model provides a best fitting model and the independence model a very poorly fitting model. Some of the other fit indexes make use of these endpoints.

Other Measures of Fit

χ^2 seems to fill our need for assessing model fit: if it is not statistically significant, we have evidence that our model may explain reality, and if it is statistically significant, our model does not explain the data. Why do we need other fit indexes? Unfortunately, χ^2 has some problems as a measure of fit. First, χ^2 is related to sample size; indeed, χ^2 is calculated as $N - 1$ times the minimum value of the fit function (FMIN on the Amos output). Thus, given the same matrix and a sample size of 10,000 instead of 1000, the χ^2 would be approximately 10 times larger than the current value of 2.166. A χ^2 of 21.66 (actually 21.68, because $N - 1$ rather than N is used in the calculations), again with 2 *df*, will be statistically significant ($p < .001$), and thus we would reach the conclusion that the model did not fit the data, an opposite conclusion from the one we reached with the sample size of 1000. (A reminder: with SEM the *df* depend on the number of model constraints, not the sample size.) This weakness of χ^2 is one reason alternative measures of fit have been developed. Most SEM users report χ^2 but also report other fit statistics as well.

Several fit indexes compare the fit of the existing model with that of the null, or independence model. The comparative fit index (CFI) and the Tucker–Lewis index (TLI, also known as the nonnormed fit index, or NNFI) are two common such indexes. The CFI provides a population estimate of the improvement in fit over the null model (although null models are the most common comparison, the CFI can also be calculated with more restricted but

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	13	2.166	2	.338	1.083
Saturated model	15	.000	0		
Independence model	5	817.868	10	.000	81.787

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.008	.999	.994	.133
Saturated model	.000	1.000		
Independence model	1.998	.715	.572	.477

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.997	.987	1.000	.999	1.000
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.200	.199	.200
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Default model	.166	.000	8.213
Saturated model	.000	.000	.000
Independence model	807.868	717.735	905.396

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	.002	.000	.000	.008
Saturated model	.000	.000	.000	.000
Independence model	.819	.809	.718	.906

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.009	.000	.064	.854
Independence model	.284	.268	.301	.000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	28.166	28.324	91.967	104.967
Saturated model	30.000	30.181	103.616	118.616
Independence model	827.868	827.929	852.407	857.407

Figure 13.11 Fit indexes for the Homework model.

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	.028	.028	.036	.028
Saturated model	.030	.030	.030	.030
Independence model	.829	.738	.926	.829

HOELTER

Model	HOELTER .05	HOELTER .01
Default model	2763	4248
Independence model	23	29

Figure 13.11 (Continued)

substantively meaningful models). The TLI provides a slight adjustment for parsimony and is relatively independent of sample size (Tanaka, 1993). For both indexes, values approaching 1.0 suggest a better fit; common rules of thumb suggest that values over .95 represent a good fit of the model to the data, and values over .90 represent an adequate fit (cf. Hayduk, 1996, p. 219; Hu & Bentler, 1999).

Another problem with χ^2 and its associated probability is that p is the probability that a model fits perfectly in the population, even though most researchers argue that a model is designed only to approximate reality. The root mean square error of approximation (RMSEA) is designed to assess the *approximate* fit of a model and may thus provide a more reasonable standard for evaluating models. RMSEAs below .05 suggest a "close fit of the model in relation to the degrees of freedom" (Browne & Cudeck, 1993, p. 144), in other words a good approximation. Browne and Cudeck further speculated that models with RMSEAs below .08 represented a reasonable fit, with those above .10 representing a poor fit. Research with the RMSEA supports these rules of thumb (i.e., values below .05 suggesting a good fit; Hu & Bentler, 1999), as well as its use as an overall measure of model fit (Fan, Thompson, & Wang, 1999). Other advantages of RMSEA include the ability to calculate confidence intervals around RMSEA, the ability to use RMSEA "to test a null hypothesis of *poor* fit" (Loehlin, 2004, p. 69), and the ability to conduct power calculations using RMSEA (MacCallum, Browne, & Sugawara, 1996). Conceptually, you can think of RMSEA as representing the degree of misfit per degree of freedom.

One final, useful measure of fit (or misfit) is the standardized root mean square residual (SRMR). We approached the topic of fit by discussing the difference between the actual covariance matrix used to estimate a model and the covariance matrix implied by the model. If you average these differences, you get the root mean square residual. (To keep the negative values from canceling out the positive values, you'd need to first square the values and then take the square root of the final average number.) The SRMR is the standardized version of the root mean square residual. Because correlations are standardized versions of covariances, the SRMR is conceptually equivalent to the average difference between the actual correlations among measured variables and those predicted by the model. Hu and Bentler's (1998, 1999) simulation research suggests SRMR as among the best of the fit indexes, with values below about .08 suggesting a good fit of the model to the data (this value may be a little high; .06 may be a more reasonable value for SRMR). The SRMR is not produced automatically in Amos but is easily obtained (select the "Plugins" menu, then Standardized RMR).

I currently use RMSEA for my primary measures of the fit of a single model, supplemented by SRMR and CFI or TLI, or sometimes other indexes. As we will soon see, it is also possible, indeed desirable, to compare the fit of competing models; we will use different fit indexes for this purpose. Note, however, that thinking and research about fit indexes are in a constant state of development. The advice I (or others) give as this is written is different from what I would have given 5 years ago and may well be different from what I will advise 5 years in the future. Because of this state of flux, and because much advice about fit indexes is based on the experience of the user, my advice may also be different from that of others. See the section at the end of this chapter for additional thoughts concerning fit indexes.

Focus again on Figure 13.11. The RMSEA for our Homework model was .009, with a 90% confidence interval of .000 to .064 (Lo 90 to Hi 90 in the figure). The CFI and TLI were 1.0 and .999, respectively. Although not shown in the figure, the SRMR for this model was .0085, suggesting an average difference between the actual and predicted correlations of only .0085. All indexes suggest a good fit of the model to the data; the model could indeed have generated the data.

Comparing Competing Models

Another major advantage of SEM programs is that we can use them to compare competing theoretical models (and the hypotheses embedded in these competing models) via the fit statistics. An example will illustrate.

Suppose in your reading of the literature on the effects of homework you came across evidence that homework and school learning are unrelated. Perhaps the evidence is in the form of research that suggests that homework has no real effects on achievement or grades. Or perhaps the evidence is in the form of informal theory that suggests that homework really should not affect learning, or vice versa. Whichever is the case, we could test these hypotheses by comparing models embodying them with our initial model (Figures 13.7 and 13.8). One such model will delete the paths from Previous Achievement to Homework and from Homework to Grades. This model asserts that students' previous achievement has no effect on the time they spend working on homework, and such time spent on homework also has no effect on students' grades. Stated differently, this model embodies the hypothesis that homework is unrelated to academic performance, either as an effect (the path from Previous Achievement to Homework) or as an influence (the path from Homework to Grades).

The standardized results of this model are shown in Figure 13.12, which also shows some of the relevant fit indexes. We will focus primarily on the RMSEA (.128), which suggests a poor fit of this model to the data. This assessment is supported by the TLI of .797 and the statistically significant χ^2 of 69.61 with 4 degrees of freedom. The CFI (.919) and the SRMR (.071), in contrast, suggest a so-so model. The model is, among other things, a good illustration that the various fit statistics often present different pictures and lead to different conclusions if used in isolation. Nevertheless, with a primary focus on RMSEA, we conclude that this model does not fit the data well, and we will likely reject the model as a good explanation of the relations between homework and learning.

We can address our primary questions more directly, however, by comparing the results of this model with the results of our initial model. That model fit well, whereas this model did not; but are the *differences* between the fits of the two models meaningful or statistically significant? We can use the fit indexes to make these comparisons, as well. Interestingly, although χ^2 has problems as a measure of fit of a single model (what I will henceforth call a "stand-alone" measure of fit), it often works well for comparing competing models (Loehlin, 2004). Furthermore, if the models are nested (meaning that one can be derived from the other by deleting paths), this comparison can be statistical rather than qualitative.

When two models are nested, the more parsimonious model (the model with fewer free, or estimated, parameters) will have a higher *df* (recall that *df* is a measure of parsimony) and a larger χ^2. The χ^2 and *df* for the less parsimonious model can be subtracted from those of

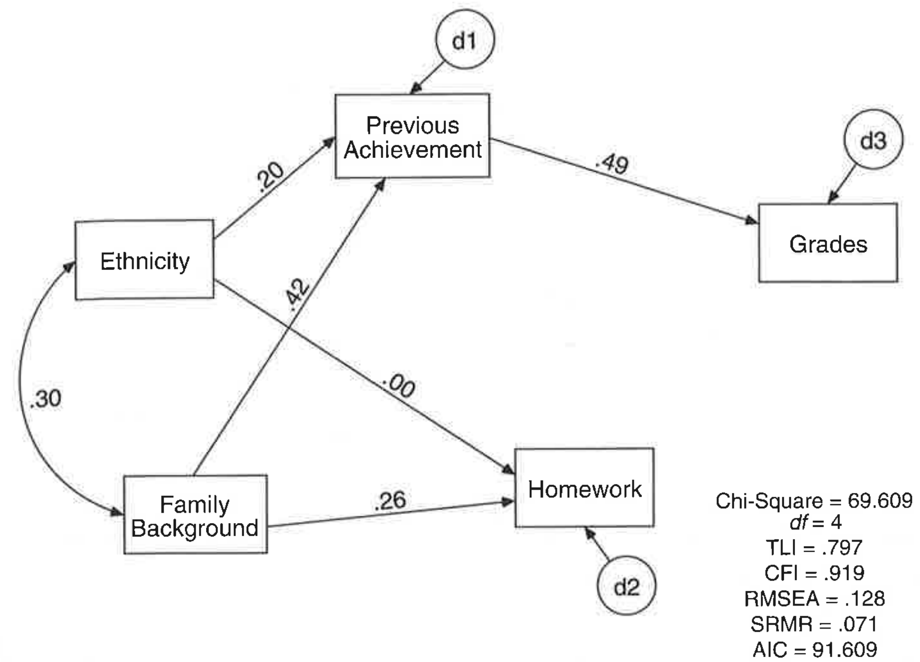

Figure 13.12 Does Homework indeed affect Grades? Compare the fit of this model with the earlier Homework model.

the more parsimonious. The resulting change in χ^2 ($\Delta\chi^2$) is also a χ^2 distribution and may be compared to the change in *df* for the two models. Again, models are nested when one can be derived from the other by deleting paths or correlations. This second model—the one with one or more paths deleted and the higher *df*—will be a subset of the first and nested within the first model.

The model shown in Figure 13.12 (no-homework-effects model) is a more parsimonious, more constrained version of the model shown in Figure 13.7 (the initial model); two paths that were estimated in the initial model were constrained to zero in the no-homework-effects model. This model is nested within the initial model. The no-homework-effects model had a χ^2 of 69.609, with 4 *df*. We subtract the corresponding values for the initial model (χ^2 = 2.166, $df = 2$) from those for the no-homework-effects model to obtain a $\Delta\chi^2$ of 67.443, with a Δ *df* of 2. If you look up these values in the probability calculator or spreadsheet,[4] you will find an associated probability of < .001; the additional constraints on the no-homework-effects model resulted in a statistically significant increase in $\Delta\chi^2$.

This finding, that the $\Delta\chi^2$ is statistically significant, means that not only does the no-homework-effects model fit worse than the initial model, but it fits statistically significantly worse. Although the no-homework-effects model is more parsimonious than the initial model, the parsimony comes at too great a cost in terms of model fit, and we reject these constraints on the model and stick with the initial model. What this means, in turn, is that we can reject the hypothesis that time spent on homework is unrelated to academic performance.

The process of comparing competing models can be used to test competing models and hypotheses, but it can also bolster, or undermine, our faith in our preferred models. "The fact that one model fits the data reasonably well does not mean that there could not be other, different models that fit better. At best, a given model represents a tentative explanation of

Table 13.4 Comparison of Fit Indexes for Alternative Models of the Effects of Homework on High School Students' Grades

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	*RMSEA (90% CI)*	*SRMR*	*CFI*	*AIC*
Initial	2.166	2				.009 (.000–.064)	.009	1.00	28.166
No Homework Effects	69.609	4	67.443	2	<.001	.128 (.103–.155)	.071	.919	91.609
Background Effect	1.329	1	.837	1	.360	.018 (.000–.089)	.008	1.00	29.329

the data. The confidence with which one accepts such an explanation depends, in part, on whether other, rival explanations have been tested and found wanting" (Loehlin, 2004, p. 61).

We can also use $\Delta\chi^2$ to test the assumptions we made when we developed our initial model. Recall that we assumed that Ethnicity and Family Background had no direct effect on students' Grades, but that their effects were indirect through Previous Achievement and Homework. We could test whether these assumptions are, in fact, supported by freeing these parameters and studying the change in fit of the model. Table 13.4 shows fit statistics for the two models already discussed, plus a model labeled Background Effect, in which the path from Family Background to Grades was freed, or estimated. As you can see, this model is less parsimonious than the initial model. The $\Delta\chi^2$ for this model was .837 with 1 *df*; the $\Delta\chi^2$ is not statistically significant. In this case, the two models had nearly equivalent fit. The more relaxed (background effect) model did not fit statistically significantly better; the more parsimonious (initial) model did not fit statistically significantly worse. In other words, the models had equivalent fit. In this case, we favor the more parsimonious of the two models, the initial model. Therefore, our initial assumption that Family Background would affect Grades only through other variables was supported. (Earlier in the text I suggested that you memorize the factoid that with a reasonable sample size, a *t* of approximately 2 is statistically significant. It would also be worth remembering that with 1 *df*, a $\Delta\chi^2$ of around 3.9 is statistically significant.)

Note that we could also have evaluated the statistical significance of the path from Family Background to Grades by focusing on the CR (critical ratio, or *t*) in the Amos printout. The *t* was .915, which is not statistically significant, thus also supporting our initial assumption of the lack of direct effect of Family Background on Grades. When single parameters are tested, $\Delta\chi^2$ and *t* will usually, but not always, give the same answer. $\Delta\chi^2$ can be used to test the statistical significance of multiple changes to a model, whereas *t* focuses on only one parameter at a time.

We could have freed both paths that were constrained to zero in the initial model (Family Background to Grades and Ethnicity to Grades). In this case, the new model will be just-identified, with χ^2 and the *df* both equal to zero. Thus, the $\Delta\chi^2$ comparing this model with the initial model equals the value for the χ^2 for the initial model (2.166, $df = 2$), which was not statistically significant ($p = .338$). Perhaps this comparison makes it obvious that, strictly speaking, what we are testing with overidentified models is the overidentifying restrictions (constraints) on the model, *not* the model as a whole.

We can also use fit statistics to clean up our models. Note that the path from Ethnicity to Homework was not statistically significant in the initial homework model. One alternative model worth investigating is one in which this path is deleted. With this change, $\Delta\chi^2$ is statistically not significant; this more parsimonious model thus fits as well as does the initial model. Although it is perfectly reasonable to use $\Delta\chi^2$ and other fit statistics to clean up models, keep in mind that this process is fundamentally different from the other model comparisons we have made. Our previous model comparisons were designed to test hypotheses drawn from theory and previous research. Model modifications to remove statistically nonsignificant paths are not theoretical; instead, they are based on the data themselves. They should not be accorded the same weight as theoretically derived model modifications until they are tested against new data. If you do a lot of data-based model revisions, you should recognize that you are conducting exploratory, rather than theory testing, research.

To reiterate, our rule of thumb is that if $\Delta\chi^2$ is statistically significant it means that the more parsimonious model has a statistically significantly worse fit than does the less parsimonious model. If you use this methodology, you would then reject the more parsimonious model in favor of the less parsimonious one. If, on the other hand, the $\Delta\chi^2$ is not statistically significant, then this means that the two models fit equally well (within a reasonable margin of error). Because we value parsimony, in this case you would reject the less parsimonious model in favor of the more parsimonious one.

Table 13.4 also includes one more fit index that can be used to compare competing models. The Akaike Information Criterion (AIC) is a useful cross-validation index in that it tends to select models that would be selected if results were cross-validated to a new sample (Loehlin, 2004). Another useful feature of AIC is that it can be used to compare competing models that are not nested. Smaller values of AIC are better, and thus if we use the AIC to compare the models in Table 13.4, we will continue to favor the initial model over its competitors. Another, related measure is the Bayes Information Criterion (BIC in the Amos output); the BIC includes a stronger adjustment for parsimony than does the AIC. Another, related index that I currently use is the sample size adjusted BIC, the aBIC. Its parsimony adjustment is between that of the AIC and the BIC. The aBIC is not currently computed in Amos but is relatively easy to calculate using other fit information provided. See, for example, David Kenny's web site (http://davidakenny.net/cm/fit.htm). The Amos manual shows how to calculate the fit indices used in that program. aBIC is produced in Mplus.

The table shows the values for the RMSEA, along with its 90% confidence interval. These values can also be used to compare competing models either informally, by choosing the model with the lowest RMSEA, or more formally, by comparing the point value for one model with the 90% CI for another model. Using either approach, we will favor the initial model as being better fitting than the no-homework-effects model and more parsimonious (but equivalent fitting) compared to the background effect model. Some researchers also use the CFI to compare competing models in tests of invariance (e.g., Cheung & Rensvold, 2002; see chapter 19).

I currently use $\Delta\chi^2$ as my primary index for comparing competing models if these models are nested and given a reasonable sample size (say 150 to 1000). For nonnested models, the AIC and aBIC have worked well in my experience.

MORE COMPLEX MODELS

Equivalent and Nonequivalent Models

Equivalent Models

We saw in Chapter 12 that with just-identified models path directions could be reversed, leading to very different conclusions, without any warning that one model was correct and the other incorrect. In other words, these models (e.g., Figures 12.3 and 12.7) are equivalent;

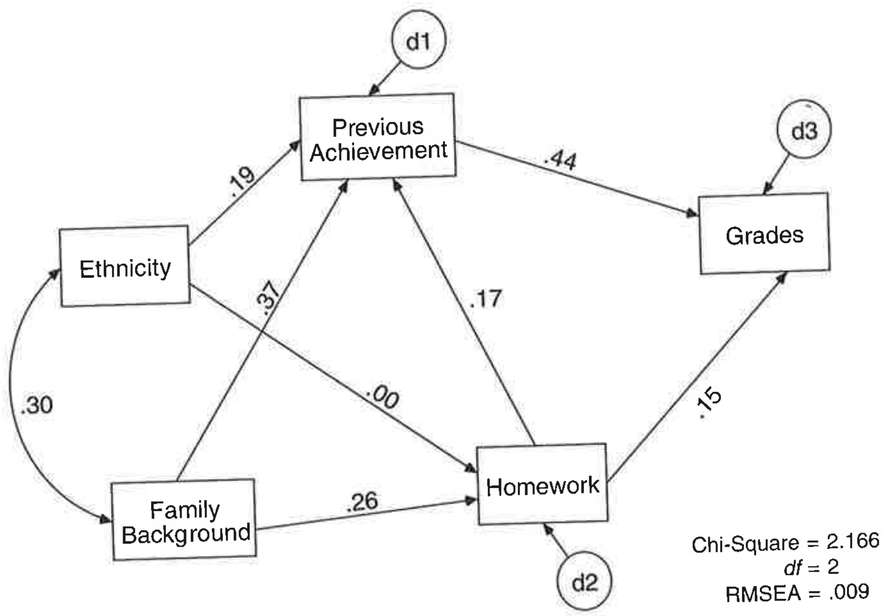

Figure 13.13 An equivalent model. Note that the Previous Achievement to Homework path is reversed, but the fit indexes are identical to those of the initial Homework model.

we cannot differentiate them by their fit. This makes sense, because just-identified models fit perfectly, and thus we cannot differentiate them by their fit.

We have seen in this chapter that one advantage of overidentified models analyzed through SEM programs is that they provide measures of fit of the model to the data. We can use these fit indexes to compare models, to reject those that fit less well and tentatively accept those with better fit. What may not be obvious is that it is also possible, in fact likely, to have equivalent overidentified models. Equivalent models are those that produce the same fit statistics as the original model and thus cannot be differentiated from that model based on fit. It is often possible to reverse a path or to replace a path with a correlation without any change in the fit of the model. For example, Figure 13.13 shows the results of an analysis in which the path from Homework to Previous Achievement was reversed (compared to Figure 13.8). As can be seen in Figure 13.13, the χ^2, *df*, and RMSEA are all the same as in the initial analyses of this model (Figure 13.11; and although not shown, the rest of the fit indexes are also identical). The two models, with the path between Homework and Previous Achievement drawn in opposite directions, are statistically equivalent and cannot be differentiated. There are, in fact, numerous equivalent models to most target models, and you should consider them as you focus on a particular model.

Stelzl (1986) and Lee and Hershberger (1990) provided rules for generating equivalent models. The main gist of these rules is summarized briefly here. For this presentation, I have assumed that the beginning models are recursive.

1. For a just-identified model, a path from *a* to *b* (symbolized as →) may be replaced by a path from *b* to *a* (→) or by a correlation between *a* and *b* (if *a* and *b* are exogenous). Endogenous variables may not have simple correlations, but their disturbances may be correlated.[5] Thus, a path from endogenous variable *a* to endogenous variable *b* may be replaced by a correlation between the disturbances of *a* and *b* (I will symbolize both types of correlations by for this ↔ discussion). All these possibilities are equivalent, meaning you can also replace ↔ with ←. This is simply another way of stating that all just-identified models are statistically equivalent because they all fit the data perfectly.
2. More importantly, for overidentified models, portions of these models may be just identified. For the just-identified portions of models, these same rules apply. That is, you can replace → by ← or by ↔ (or vice versa), and the model will be equivalent. So, for example, in Figure 13.13 the model is just-identified through the variable Homework. This is why we can reverse the Homework–Previous Achievement path and still have an equivalent model.
3. For portions of the model that are overidentified, if *a* and *b* have the same causes, → (a path from a to b) may be replaced by ↔ or by ←. Thus, for the model in Figure 13.7, reversing the path from Homework to Grades will not result in an equivalent model, because the two variables do *not* have the same causes.
4. For portions of the model that are overidentified, when *a* and *b* do *not* have the same causes, the substitutions are slightly more complex. A path from *a* to *b* may be replaced by ↔ if the causes of *b include* all the causes of *a*. You could not replace the path from Homework to Grades with a correlated disturbance between d2 and d3 because the causes of Grades do not include all the causes of Homework. Ethnicity and Family Background are influences on Homework but not Grades. In addition, correlated disturbances can be replaced by a path from *a* to *b* if *b* includes all causes of *a*.

Figure 13.14 shows several equivalent models to our original Homework model (from Figure 13.7, also shown as model A in Figure 13.14). Make sure you understand why each is equivalent to the original model. It is worth noting that these rules can be applied repeatedly, which is how the final model (model F) is derived. The derivation of each model is explained in note 6.[6]

It should be obvious from a study of Figure 13.14 that the presence of equivalent models may threaten the causal conclusions from our research. If all these models are statistically equivalent to our preferred model, how can we assert, for example, that Previous Achievement affects Homework, rather than Homework affecting Previous Achievement? I encourage you to generate and consider alternatives to your model of choice. You may discover alternatives that make as much sense as your original model, or you may begin to feel more comfortable with your initial model. It is certainly better to consider equivalent models and either revise your models or defend your reasoning prior to publication rather than after! But, in reality, the answer to the threat of equivalent models is the same as the method of devising strong models to begin with: consider logical and actual time precedence, build in relevant theory and research, and carefully consider the variables involved.

What should we do, however, when equivalent models remain plausible even after such considerations? As we will see, one possible solution is to devise *non*equivalent models that evaluate the different possibilities; another possibility is the use of longitudinal data.

The Lee and Hershberger rules apply to portions of nonrecursive models as well, but the rules presented here will cover most models of interest in this text. See Lee and Hershberger (1990) for more information; the rules are also summarized and well illustrated by Hershberger (2006) and by Kline (2011). MacCallum, Wegener, Uchino, and Fabrigar (1993) illustrated problems that arise from not considering equivalent models.

Directionality Revisited

If some overidentified models are equivalent, it follows that some overidentified models are not equivalent and that we can use the same rules to generate nonequivalent models. These, in turn, may help us deal with one problem we encountered with simple just-identified models:

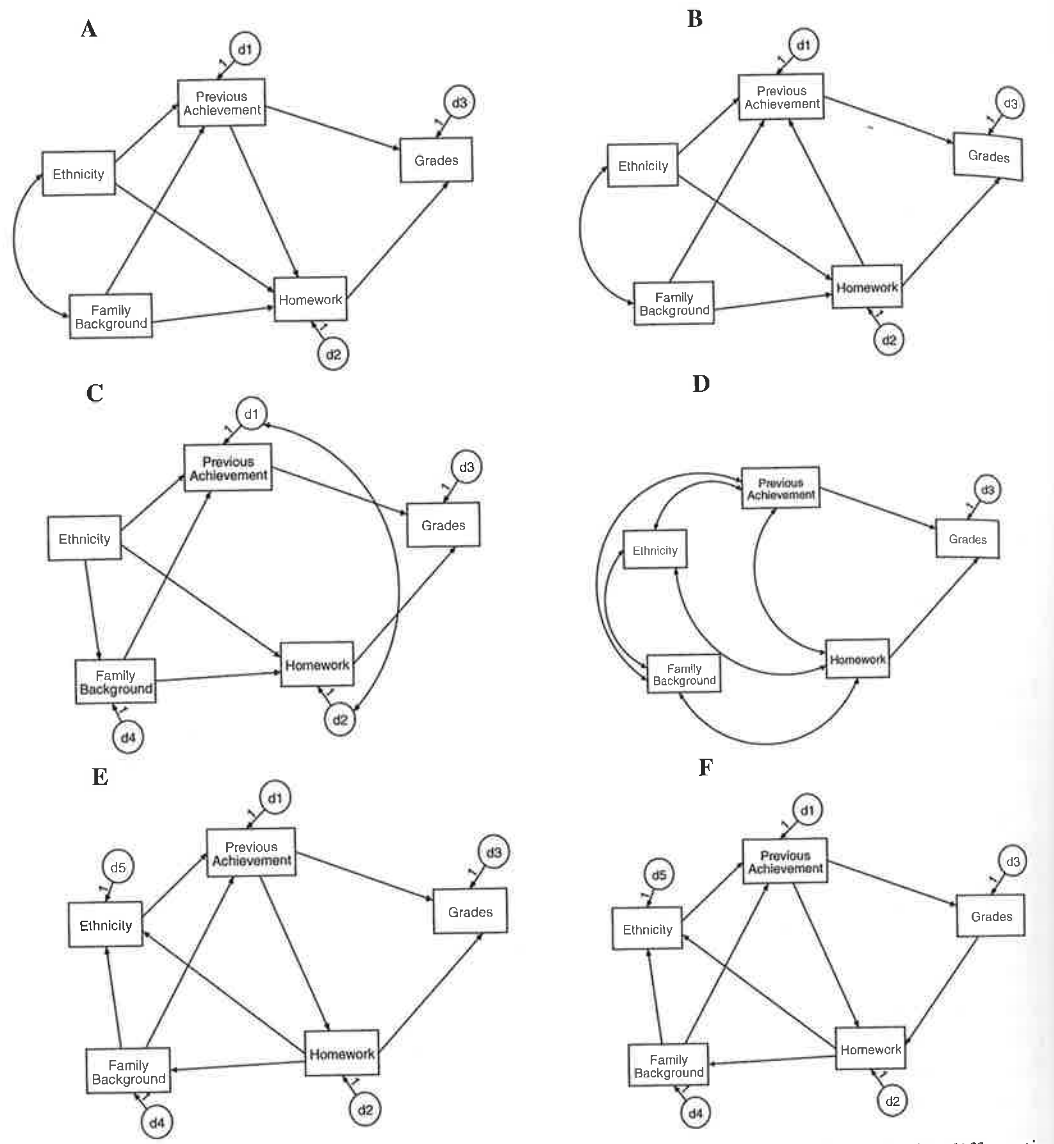

Figure 13.14 Equivalent models. All the models are equivalent to Model A and cannot be differentiated from it based on fit.

uncertainty concerning causal direction. As you will see, although the problem of equivalent models is a danger to SEM interpretation, an understanding of the rules of equivalent models can lead to the development and testing of nonequivalent models, which can be a blessing.

Figure 13.15 shows one more version of the homework model, one in which the path from Homework to Grades is reversed, replaced by a path from Grades to Homework. This direction does not make sense based on time precedence (Homework includes information from 8th and 10th grades, whereas Grades are from 10th grade). Still, as demonstrated in Chapter 12, if we estimate a just-identified version of this model, there will be nothing in our analysis to tell us that it is incorrect. The current version is overidentified. More importantly, this model is not equivalent to the original model. Grades and Homework do not have the

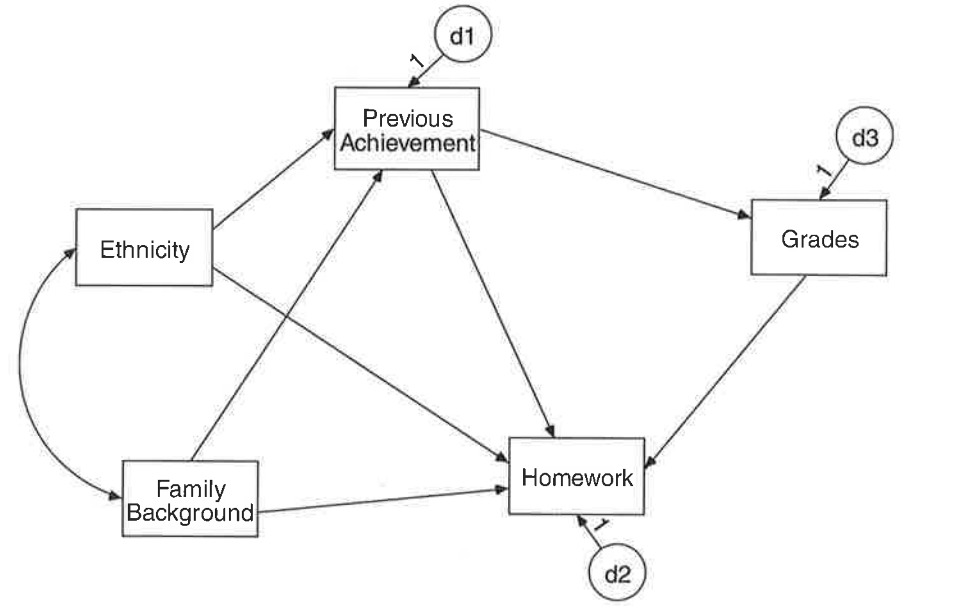

Figure 13.15 Reversing the Homework to Grades path results in a nonequivalent homework model.

same causes (rule 3), and thus the reversal of the path does not result in an equivalent model. If the models are not equivalent, does that mean that the fit indexes may help spot the error in our model? In a word, yes.

Figure 13.16 shows the solved "wrong direction" model with a few of the relevant fit indexes. Note that if we look at the RMSEA (or other stand-alone fit indexes), this model will be deemed acceptable. Of more interest, however, is to compare this model with the initial "correct" homework model. We can't use $\Delta\chi^2$ because the two models are not nested; you cannot arrive at one by deleting paths from the other. Indeed, the models are equally parsimonious (they have the same *df*). We can still use the AIC to compare nonnested models, however. As you can see, if you compare the AIC from Figure 13.16 with the fit indexes for the original model (shown in Figure 13.11), the AIC for the original model is smaller. The rule of thumb for AIC is to favor the model with the lower value; we would thus favor the original model over the model with the Homework–Grades path drawn in the wrong direction. The judicious use of nonequivalent models may indeed help us answer nagging questions of directionality!

You may wonder why this should work. Recall the genesis of the fit indexes: a comparison of the actual correlation–covariance matrix with the matrix implied by the model. Quite simply, Figure 13.16 implies a slightly different covariance matrix than does the model shown in Figure 13.8, and the matrix implied by the model shown in Figure 13.8 comes closer to the actual matrix.

Practically, the easiest way to develop such nonequivalent models is to include variables that uniquely cause one of the variables in question. That is, include variables in the model that are influences of the presumed cause but not the presumed effect and variables that are influences on the presumed effect but not the presumed cause. In other words, include some relevant *noncommon* causes in the model. Thus, although we saw in Chapter 12 that noncommon causes are not *required* for the model to be valid, we now see they may help in dealing with other problems. Likewise, intervening (mediating) variables can help in the development of nonequivalent models and thus may be valuable for this purpose as well.

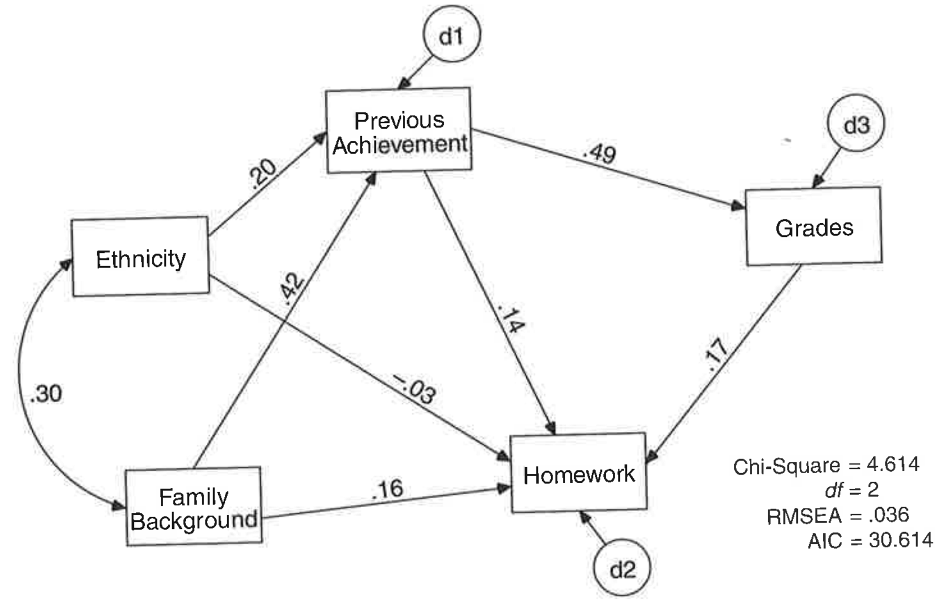

Figure 13.16 The nonequivalent homework model demonstrated a worse fit to the data.

Nonrecursive Models

Another advantage of SEM programs is that they can be used to analyze nonrecursive models, or models with feedback loops. Suppose you were interested in the influences on partners' levels of trust in marriage and other close male–female relationships. It makes sense that my level of trust in my wife may be affected, in part, by my own personal and psychological characteristics. My trust may also be affected by my wife's level of trust in me, however, and vice versa. If I trust my wife more, she will likely trust me more, and so on. Trust likely has reciprocal effects. Your theoretical model might look something like that shown in Figure 13.17. The model posits that one's trust in his or her partner is affected by one's own characteristics (self-esteem and perception of the partner's desire for control), as well as by the partner's own level of trust. This model is a smaller version of one posited and tested by John Butler (2001).

Recall that the tracing rule does not work with nonrecursive models but that we can develop formulas for the paths using the first law of path analysis. If you develop equations for the model shown in Figure 13.17, you find that, unlike recursive models, the formulas no longer are equivalent to those for regression coefficients from multiple regression. This is simply a convoluted way of saying that with nonrecursive models you cannot use ordinary multiple regression to estimate the paths.

It is possible, however, to use SEM programs to estimate models such as those shown in Figure 13.17. Some results of such an analysis are shown in Figure 13.18; they suggest that each partner's trust is indeed affected by the other's trust. Self-Esteem had a positive effect on Trust, and Perception of Control had a negative effect, although the relative magnitudes of these effects were different for men and women. You will have a chance to return to this model in the exercises. (The data that produced these results are simulated because the original article did not include the correlation or covariance matrix. These simulated findings are consistent with those of the original article, however.)

I have presented this model as an example of the use of nonrecursive models to answer questions in which we expect there to be reciprocal effects. These are common in analyses

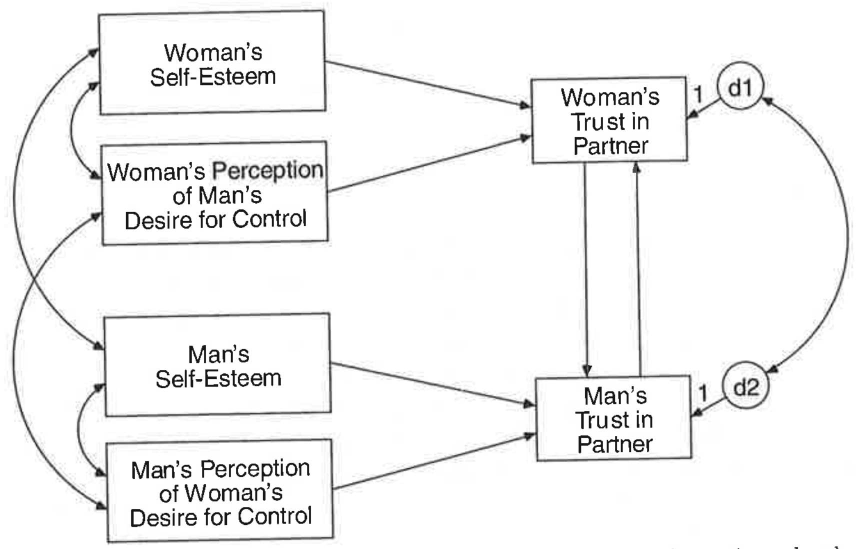

Figure 13.17 Nonrecursive model to test the reciprocal effects of partners' trust in each other.

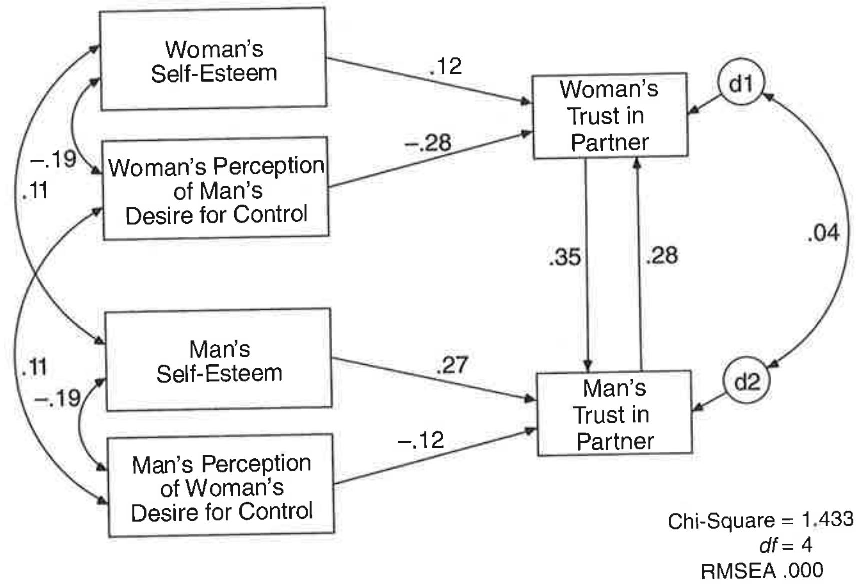

Figure 13.18 Standardized solution, partner trust model. The data are simulated, but based on research reported by Butler (2001).

of data from couples or other pairs of people. One of the best-known nonrecursive models, extensively analyzed and used as an example in many SEM manuals, was devised by sociologist Otis Dudley Duncan and colleagues to estimate the effects of friends on each other's occupational and educational aspirations (Duncan, Haller, & Portes, 1971). As you might expect, nonrecursive models are also used to settle questions of causal sequence (e.g., Reibstein, Lovelock, & Dobson, 1980).

Nonrecursive models are considerably more complex than this simple overview, however, and are beyond the scope of this book. If you are interested in pursuing nonrecursive models, you will need to study such models in considerably more depth. Kline (2006) and Loehlin (2004) provide a more detailed introduction, Rigdon (1995) presents a detailed discussion of identification issues for nonrecursive models, and Hayduk (1996) presents interesting issues related to nonrecursive models.

Longitudinal Models

Another method of answering questions about the reciprocal effects of variables on one another is through longitudinal models. Indeed, if you focus on our homework models, you will see that they take advantage of this technique. These models focus on the effects of homework on learning in later grades (subsequent GPA), while controlling for achievement in an earlier grade (Previous Achievement in 8th grade).

Do job stress and emotional exhaustion (or burnout) have reciprocal effects? Figure 13.19 shows a longitudinal model designed to answer this question for physicians surveyed in the United Kingdom (McManus, Winder, & Gordon, 2002). The physicians were surveyed in 1997 and again in 2000; the variables in the model should be self-explanatory. The data (stress burnout longitudinal.xls) and this model are on the Web site (stress burnout longitudinal 5.amw).

The model is barely overidentified (with 1 *df*); there is no path from Personal Accomplishment at time 1 to Stress at time 2. The results suggest that Stress and Emotional Exhaustion indeed have reciprocal effects. Stress increases Exhaustion, which, in turn, increases subsequent Stress. It is worthwhile to compare this model to one in which it is assumed that Stress affects future Stress only via the indirect effect through Exhaustion (full mediation).

Longitudinal models can also help bolster the reasoning behind the paths we draw, even in the presence of equivalent models. If Emotional Exhaustion is measured in 1997 and Stress in 2000, it is easier to argue that the proper direction is from Exhaustion to Stress than if they are measured concurrently. Still, I don't want to oversell the ability of nonrecursive and longitudinal models to answer questions about the direction of influence; the results are

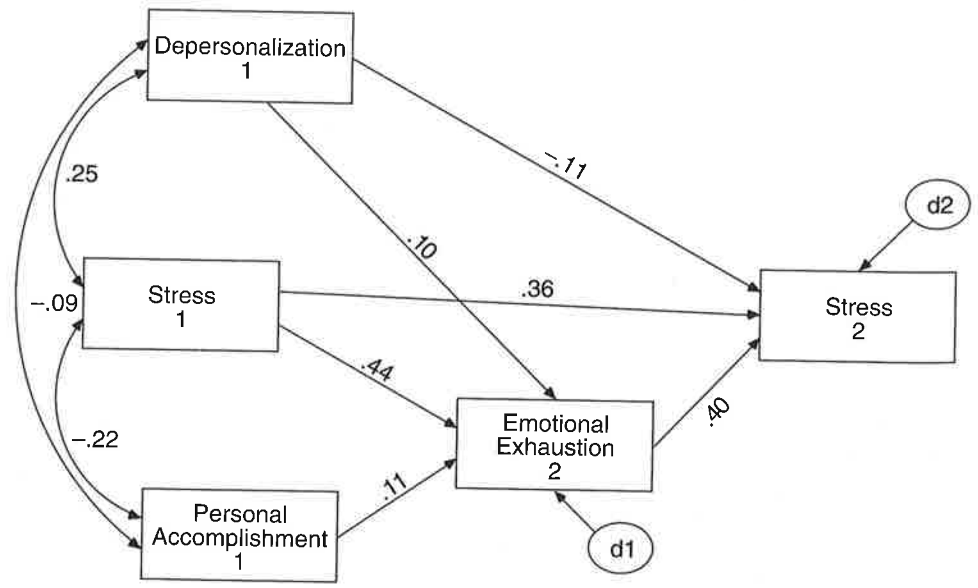

Figure 13.19 Reciprocal effects of Stress and Emotional Exhaustion, estimated via longitudinal data. The model is based on research with physicians (McManus et al., 2002).

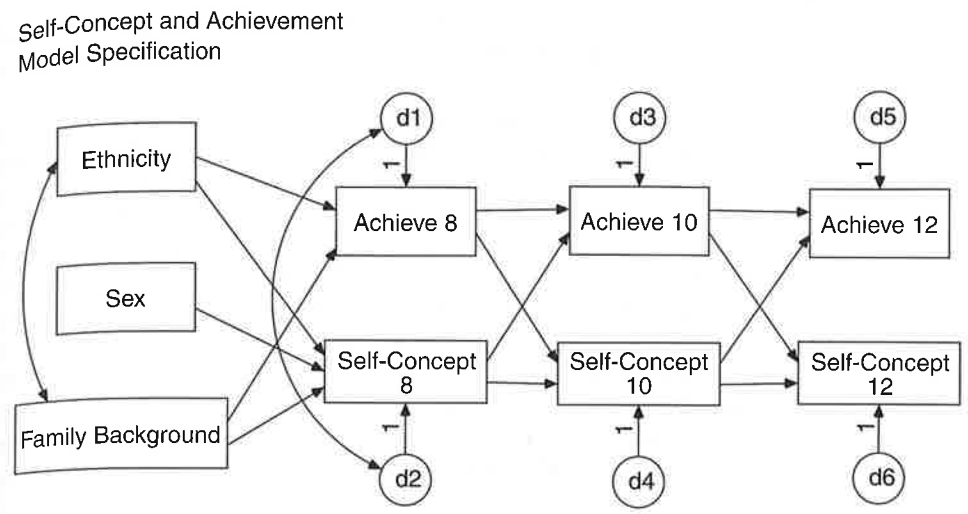

Figure 13.20 Potential longitudinal panel model designed to determine the extent of the effect of self-concept on achievement, and vice versa.

not always as clear as we would like them to be. I have provided fairly clean and clear-cut examples here to illustrate the possibilities.

Figure 13.20 shows a special type of longitudinal model known as a panel model. A panel model has the same set of two or more variables measured two or more times. It is often used to test questions of reciprocal causation or settle issues of causal predominance. The model shown could be tested with the NELS full data (including the 12th-grade data, not included in our NELS subsample). Note that the Achievement tests (the same or similar tests) and the self-concept measure are administered three times; the model, as shown, has 21 df. If the results of the analysis showed a substantial effect from Achievement at every time period to Self-Concept at the next but not the reverse (Self-Concept not affecting Achievement) then we would feel more comfortable in specifying a path from Achievement to Self-Concept in subsequent cross-sectional or longitudinal research. Note the correlated disturbances for the two variables of interest in 8th grade. In this case, the correlated disturbance may serve two purposes: it can take into account that we have not specified any effects between Achievement and Self-Concept at Grade 8 (perhaps additional correlated disturbances are needed at the other time points?), and that there may be other common causes of these variables that we have not considered. The time lag between measures in panel and other longitudinal model eases our concerns about specifying a causal direction, but keep in mind that the lag needs to be long enough for the causal process to have worked. For more information about longitudinal models, in general, and panel models, in particular, see Little, 2013.

ADVICE: MR VERSUS SEM PROGRAMS

We have seen that with just-identified models SEM programs provide the same information for a path analysis as we get with multiple regression programs. With overidentified models, however, there are advantages in using SEM programs. If you have a choice, which should you use? Here's my advice:

1. If you plan to analyze a single, just-identified recursive model, either MR or a dedicated SEM program will work just fine.

2. If you plan to analyze an overidentified model or compare several competing models, use a SEM program. If you plan to analyze a nonrecursive model, use a SEM program.
3. If you are using a MR program to conduct a path analysis, there is no real benefit in specifying overidentified models. Instead, what I suggest is a more qualitative evaluation of fit. By this I mean that prior to analysis you should try to predict, based on previous research and theory, which paths will be close to zero, which should be large, which should be positive, which should be negative, and so on. I'm not suggesting that you necessarily need to make these as formal predictions, but you should spend some time thinking about what you expect each path to look like. After conducting the analysis, see how your predictions fared. If the paths you thought should be close to zero were, in fact, close to zero, and so on, you can have much more faith that your model may be a faithful approximation of the way the phenomenon you are studying actually works. If, on the other hand, many of your predictions were wrong, you should be more cautious in your interpretation and should rethink you model and double-check your analyses.
4. If you are using a SEM program to conduct a path analysis, it is worthwhile to try to specify overidentified models rather than just-identified models. Again, spend some time comparing your model to what you know based on theory and previous research. Are there paths that you can set to zero based on such information? If so, delete them from your model (you can always test these no-effect hypotheses in subsequent models). Again, it is preferable to specify these no-effect hypotheses prior to analyzing the data, rather than after running a just-identified model and noting which paths were statistically nonsignificant. If you are using a SEM program, you should also consider the substantive hypotheses you can test by comparing competing models.

ADVICE: MEASURES OF FIT

If you are using a SEM program to conduct path analysis (or CFA or latent variable SEM) you should strive for overidentified models and use the fit information to evaluate the models and to compare competing models. It is with some trepidation that I write this section attempting to consolidate and expand my earlier advice on fit indices. Quite simply, given the number of model characteristics (e.g., sample size, number of variables, degrees of freedom, model misspecification, and unique variances) my advice will often be wrong. But if you are a beginner, you need to start somewhere. Please also see the caveats at the end of this section.

Evaluating a Single Model

As noted earlier in the chapter, I have found RMSEA, SRMR, CFI, and TLI useful for evaluating the fit of a single model, or what I've called useful "stand-alone" fit indexes. Common criteria for these fit indices are shown in Table 13.5. As noted earlier, these criteria have been generally supported in simulation studies (e.g., Hu & Bentler, 1998, 1999). These authors (Hu & Bentler, 1999) recommended using them in combination, such as SRMR and CFI. But things are not that simple. More recent research, however, has shown potential problems with cut-off criteria for good versus poor fitting models (Chen, Curran, Bollen, Kirby, & Paxton, 2008; Fan & Sivo, 2007; Marsh, Hau, & Wen, 2004). Many things affect fit indices, so adherence to rigid cutoff criteria simply will not work. For example, concerning RMSEA, "The authors' analyses suggest that to achieve a certain level of power or Type I error rate, the choice of cutoff values depends on model specification, degrees of

freedom, and sample size" (Chen et al., p. 462). I continue to use the criteria listed in the Table but not as a fixed good model/bad model criteria. If all the fit indices look good, I'm tentatively OK with a model. If some are good and some are not so good, I try to understand why and investigate how the model could be improved. Loehlin likened this approach of multiple fit indexes to having multiple watches, each of which may tell you a different time (Loehlin, 2004, chap. 2). If the watches are fairly close to one another, you will have a pretty good idea of the correct time. If they tell you vastly different times, you'd better investigate further.

I have mixed feelings about the use of χ^2 as a primary measure of fit for a single model; you should realize that other writers are more supportive. Kline, for example, suggests always reporting χ^2 and its associated *df* and *p*, and for models with a statistically significant χ^2, carefully examining the residual correlation matrix for the sources of misfit. I am less enamored with χ^2, but that may be because most of my research involves large samples (thousands of cases) and, given that χ^2 is so affected by sample size, my χ^2s are usually statistically significant. But this is not bad advice, especially if you use sample sizes in the 75 to 200 or maybe even 400 range (http://davidakenny.net/cm/fit.htm, retrieved April 1, 2014). With larger samples, I think χ^2 is less useful as a stand-alone measure of fit. And the more general point here is even more useful: when fit, as measured by your preferred indexes, is less than stellar, then you should investigate further. The residual correlations and the standardized residuals (covariances) are an excellent resource for doing so. The modification indices (and associated expected parameter change) are also useful (Heene, Hilbert, Freudenthaler, & Bühner, 2012); these will be discussed in subsequent chapters.

This difference highlights several important points. First, different writers will emphasize different measures of fit and will give different advice. Second, knowledge about the performance of various fit indexes will increase over time, and common wisdom concerning fit indexes will change over time. If you are to be a responsible user of SEM for research, you need to say attuned to new developments. I've already noted Kenny's web pages as a good source of current advice; presumably he will continue to update his advice. You should also pay attention to the conventions and norms in your own area of research, because these will differ from one area to another. Third, you should always keep in mind that even when a model fits the data well, that does not mean that the model is correct, and that you have found "truth." There may be alternative models with equivalent or better fit. And even if your model beats out all alternatives, it's just a model; it does a good job in explaining the observed relations among the variables you have looked at, and those are just a small slice of the infinite number of variables you could have considered. Fourth, and very importantly, what you should NOT do is cherry-pick your fit index to support the model that you prefer. Although you can change your preferences for fit indices over time, that change should be based on knowledge and experience, not the desire to support a particular model.

Comparing Competing Models

This lack of concrete, universally accepted rules of thumb concerning what constitutes a good model, and the fact that good models are not "correct" models, highlights a fourth important point: although stand-alone measures of fit are very useful, it is even better when we can compare the fit of alternative, competing models. As already noted, I've found $\Delta\chi^2$ useful for this purpose, when those models are nested, and given reasonable sample sizes (say up to 750 or 1,000 or so). Also useful are the AIC and other information criteria indexes, and these have the advantage of being usable and useful when models are non-nested. The various information criteria indices (AIC, BIC, aBIC) give different rewards for parsimony

(Mulaik, 2009 has shown that these parsimony "rewards" depend on sample size, and disappear with large samples). At least in my recent research I have found the aBIC to provide a happy medium between too strict versus too forgiving. All of these indexes (AIC, BIC, aBIC) are only useful for comparing competing models (they are not used or useful as stand-alone indexes), and smaller is better.

Finally, please recognize that my term "stand-alone" fit index is not common. I think it makes sense to talk about stand-alone measures of fit versus measures useful for comparing competing models, but this is not common usage. More commonly, writers will refer to measures such as CFI and TLI as incremental or relative fit indexes (because they compare the target model with a null model), and measures such as RMSEA, SRMR, sometimes and χ^2 as absolute fit indexes. Additional categorizations vary from writer to writer. Kenny adds the term "comparative fit" indexes for indexes such as AIC that are only useful for comparing competing models (http://davidakenny.net/cm/fit.htm, retrieved April 1, 2014), others refer to AIC and related measures as information-theoretic measures (e.g., Arbuckle, 2013), and so on.

Table 13.5 shows the fit indices we have discussed so far, and their usefulness (in my opinion) for evaluating a single model or competing models. Some other indices are included as well.

Table 13.5

Fit Index	*May be useful for & other notes*	*Common criteria*
χ^2	Useful as stand-alone measure with $N = 75$ to 400. Tested for statistical significance with *df*	non-significance supports the model
$\Delta\chi^2$	Comparing competing, nested models, $N \leq 1000$	Non-significance supports the model with larger *df*; significance supports the model with smaller *df*
RMSEA	Stand-alone measure of fit. Can calculate confidence intervals around RMSEA, and test whether an obtained RSMEA is statistically significantly different from some value (e.g., .05)	$\leq .05$ = good fit (close fit) $\leq .08$ = adequate fit $\geq .10$ = poor fit
SRMR	Stand-alone measure of fit. Intuitively appealing	$\leq .08$ = good fit, although $\leq .06$ may be a better criterion
CFI	Stand-alone measure of fit. Some research suggests ΔCFI may be useful in invariance testing (see Chapter 19)	$\geq .95$ = good fit $\geq .90$ = adequate fit
TLI	Stand-alone measure of fit	$\geq .95$ = good fit $\geq .90$ = adequate fit
AIC	Comparing competing models—nested or non-nested	Smaller is better
BIC	Same as AIC, but larger reward for parsimony	Smaller is better
aBIC	Same as AIC, in between AIC and BIC in reward for parsimony	Smaller is better

SUMMARY

We covered a great deal of ground in this chapter; a review is needed. In this chapter we made the transition from estimating path models using multiple regression analysis to estimating these models with programs specifically designed for structural equation modeling (SEM). Several such programs are available, each with its own advantages. Many programs have student, or demonstration, versions available, downloadable from the Web; these student versions work the same as do the full-featured programs, but generally limit the number of variables that can be analyzed. There are SEM modules available for R, the free statistical programming language. I have used the Analysis of Moment Structures (Amos) program to illustrate SEM programs. The illustrations and explanations should translate easily to other SEM programs, and the web site illustrates input and output from several SEM programs.

All our previous discussions of path analysis translate directly to path analysis via SEM programs. To illustrate, we re-estimated the parent involvement path model from Chapter 12 using Amos. One advantage of Amos is that a drawing of a path model is used as the specification of the model, and the drawing, along with the data, is sufficient for conducting the analysis. The input drawing for reanalysis of the parent involvement example was similar to the conventions we have used previously for developing path models. The one difference was that, by convention, we set the paths from the disturbances to the endogenous variables to 1, which allowed us to estimate the variance of the disturbance. (In multiple regression the variance of the disturbance was assumed to be 1, but the path was estimated.) We will follow this convention with other unmeasured–latent variables as well: setting the scale of the unmeasured variable by setting the path from it to one measured variable to 1; this convention merely says the scale of the unmeasured variable is the same as that of the measured variable.

Output from the SEM program (in this case Amos) included standardized and unstandardized path models, as well as detailed output. The more detailed output included standard errors of unstandardized coefficients and their associated t (or z) statistics, as well as tables of direct, indirect, and total effects.

Our next example was an overidentified model designed to determine the extent of the influence of Homework time on high school Grades. The model did not include all the paths that could have been drawn, a specification that is the same as drawing the paths but constraining them to a value of zero. The solved model suggested that Homework had a moderate effect on Grades, and Previous Achievement and Family Background had moderate to strong effects on time spent on Homework.

In earlier chapters I noted that overidentified models can be used to provide feedback about the adequacy of the model. A chief advantage of SEM programs is that they naturally provide such feedback. We can solve for paths using covariances, but we can also do the reverse: solve for the covariances using the solved path model. When models are overidentified, these two matrices (the actual and the implied covariance matrices) will differ to some degree. Fit statistics or indexes describe this degree of similarity or dissimilarity and provide feedback as to the adequacy of the model in explaining the data.

The degrees of freedom for a model describe the extent to which it is overidentified, or the parsimony of the model. The Homework model had 2 degrees of freedom; there were two paths we could have drawn but did not. The more we constrain values in the model to zero (or some other value), the more parsimonious the model and the larger its degrees of freedom.

Numerous fit indexes are provided by SEM programs. We focused on the root mean square error of approximation (RMSEA) as a primary index of fit for a single model; RMSEAs of .05 or less suggest a good fit, with values of .08 or less suggesting an adequate fit (cf. Browne & Cudeck, 1993). I also discussed using the comparative fit index (CFI), and the Tucker–Lewis index (TLI) as methods of assessing the fit of a single model. For these indexes, values above .95 suggest a good fit, and values above .90 suggest an adequate fit. The standardized root mean square residual (SRMR) is an intuitively appealing index of fit, and represents the average difference between the actual correlations among measured variables and those predicted by the model; SRMR values below .08 (or perhaps .06) represent a good fit. χ^2, along with the *df* and its associated probability, may be used to assess the fit of a model, with statistically significant values suggesting a lack of fit and statistically not significant values suggesting a good fit of the model to the data. Although common, χ^2 has problems as a measure of the fit of a single model.

A major advantage of SEM programs and measures of fit is that they may be used to compare competing theoretical models. We compared the fit of the initial Homework model to several competing models; these comparisons tested basic hypotheses embodied in these models. Although I downplayed the use of χ^2 as the measure of fit of a single model, I argued that if models are nested (one is a more constrained version of the other) χ^2 can be a useful method of comparing the two models. The more parsimonious model (the one with the larger *df*) will also have a larger χ^2. If the change in χ^2 is statistically significant compared to the change in *df*, our rule of thumb is to prefer the less parsimonious model; but if the $\Delta\chi^2$ is statistically not significant, our preference is for the more parsimonious model. A $\Delta\chi^2$ of close to 4 is statistically significant with 1 *df*. Other fit indexes for comparing competing models are the Akaike Information Criterion (AIC) and the sample size adjusted Bayes Information Criterion (aBIC), in which smaller values are better.

Although overidentified models allow us to compare competing models, representing competing hypotheses about the effects of variables on each other, there may be several or many models that are equivalent to our preferred model. These equivalent models may also represent competing hypotheses about effects but are statistically indistinguishable from our preferred model. I briefly explained and illustrated the rules for generating equivalent models, and noted that you should consider such equivalent models as you develop your own models. You can guard against the threat represented by equivalent models in the same way you build valid models in the first place, through careful consideration of theory, previous research, time precedence, and so on.

The flip side of equivalent models is that there are other overidentified but nonnested models that are not equivalent with the model under consideration. Such models can be very useful for testing and rejecting threats to path models. Knowing the rules for generating equivalent models also allows us to develop nonequivalent models. We illustrated this value by testing a nonequivalent version of the Homework model with the path from Homework to Grades reversed.

Other advantages of SEM programs are that they can be used to analyze nonrecursive models and can provide for more powerful analysis of longitudinal models. Longitudinal data may also be useful for overcoming some challenges posed by equivalent models by clarifying causal direction. I briefly illustrated such models but did not delve into them in detail.

We now have two methods for analyzing path models: multiple regression analysis via a generic statistical analysis program or SEM programs. If you are using MR to conduct path analysis, there is no real benefit for developing overidentified models. If you are using a SEM program, however, it is worth developing overidentified models when possible, because of the fit information the programs provide. Similarly, if you are interested in overidentified models, comparing competing models, or in more complex forms of path models, I encourage you to use a SEM program to estimate these models.

EXERCISES

1. Reproduce the Homework models used in this chapter. Make sure your results match mine (note there may be minor differences in estimates if you are using programs other than Amos). Are there additional models that you might test?
2. Try estimating a similar homework model using the NELS data.
3. In the section introducing overidentifying models, I stated that "not drawing a path is the same as drawing a path and fixing or constraining that path to a value of zero." Demonstrate the truth of this statement. Using the homework model, constrain, for example, the path from Previous Achievement to Grades to zero and check the fit of the model. Now delete that same path. Is the fit the same? Are the parameter estimates the same for the two models?
4. Focus on the equivalent models in Figure 13.14. Note the difference between these and the initial model (model A). Which rule or rules were used to produce each equivalent model? Check your answers against those in note 5. Try estimating one or two of these models to demonstrate that they are indeed equivalent.
5. Henry, Tolan, and Gorman-Smith (2001) investigated the effect of one's peers on boys' later violence and delinquency. Figure 13.21 shows one model drawn from their study, their "fully mediated" model. Family Relationships is a composite of measures of family cohesion, beliefs about family, and family structure, with high scores representing a better functioning family; the violence and delinquency variables are measures of the frequency of violent and nonviolent delinquent offenses for peers and individuals. The model is longitudinal, with Family Relationships measured at age 12, Peer variables at age 14, and Individual variables at age 17. The model is also contained in the file "henry et al.amw" on the Web site.

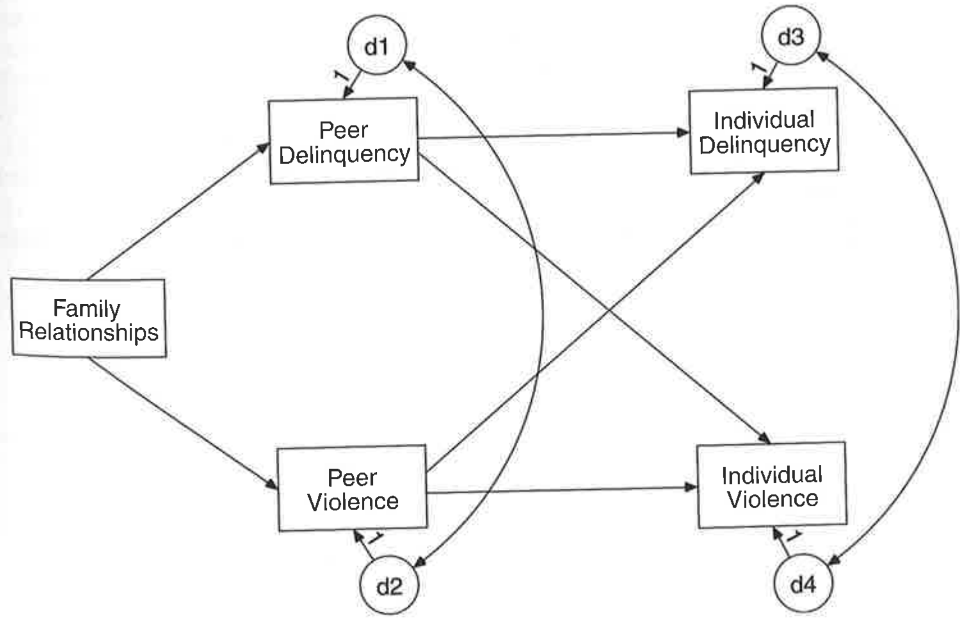

Figure 13.21 One model from Henry et al. (2001).

Data consistent with those reported in the original article are in the SPSS file "Henry et al.sav" or the Excel file "Henry et al.xls." Analyze and interpret this model. Which variable had a more important effect on boys' delinquency: peers who are delinquent or peers who are violent? Which variable was more important for boys' violence? What were the indirect effects of Family Relationships on Individual's Violence and Delinquency? Test an alternative model to determine whether Family Relationships directly affect the outcome variables.

(The Henry et al., 2001, article reported correlations among variables. The data used in this example were simulated data designed to mimic these correlations. The Family Relationships variable used here was a combination of three variables from the original article.)

6. Estimate the nonrecursive trust model from Figure 13.17. The model (trust nonrecursive model 1.amw) and the data (trust norec sim data.xls) are included on the accompanying Web site. Second, assume that the Man's Trust affects his partner, but not the reverse: delete the path from Woman's Trust to Man's Trust, along with the correlated disturbance. Are these models nested? Why? Compare the fit of the two models. What conclusions do you reach from these model comparisons?
7. Exercise 6 in Chapter 4 was "designed to explore further the nature of common causes, and what happens when non-common causes are included in a multiple regression. We will begin our analysis of these data here, and will return to them in Part 2 when we have the tools to explore them more completely."

You now have the tools to explore them more completely. To review, there are two data files for this exercise, both including variables labeled X1 X2 X3 and Y1. In both files, the three X variables are intercorrelated, but variable X2 is not a common cause of variables Y1 and X3. For the data in the first file (common cause 1.sav), variable X2 has no effect on Y1. In the second file (common cause 2.sav), variable X2 has no effect on variable X3.

Analyze these data using an SEM program. For both data sets, the model you should estimate is illustrated in Figure 13.13. Compute and examine the correlations among the variables in both data sets. All correlations are statistically significant, correct?

Now analyze the model shown for both data sets. Notice that for data set 1 the effect of X2 on Y1 is essentially zero. For data set 2, what is the effect of X2 on X3? Is X2 a common cause of X3 and Y1 in either model? Now notice the effect of X3 on Y1 in each model. What should happen to this path when the variable X2 is removed from the model?

Analyze each data set without variable X2 in the model. What happens to the magnitude of the path from X3 to Y1?

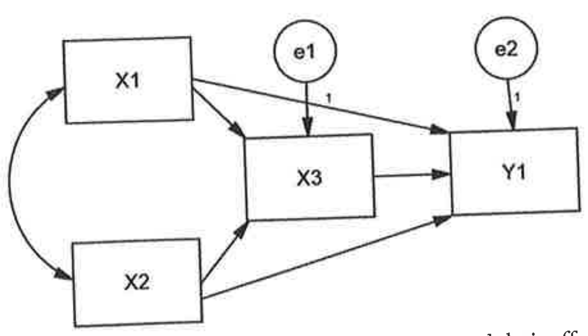

Figure 13.22 Understanding common versus non-common causes, and their effects on path estimates.

Now conduct the same analyses with the data set in "common cause 3.sav." Note the relation of X2 to X3 and Y1. Now take variable X2 out of the model. What happens to the path from X3 to Y1?

Consider what you findings mean concerning the nature of controlling for common versus non-common causes.

Notes

1 We could also analyze the NELS raw data, but would then need to consider methods of dealing with missing data in more depth than I want to right now. We will return to this issue in the chapter on latent means in SEM, Chapter 18.

2 How could you do so? It is fairly easy to do so using the tracing rules. For example, to calculate the correlation between Ethnicity and Grades implied by the model, here are the possible tracings between Ethnicity and Grades (where → represents a path and ↔ represents a correlation):

1. Ethnic → PreAch → Grades + Ethnic → PreAch → Homework → Grades + Ethnic → Homework → Grades = .20 × .44 + .20 × .22 × .15 + –.04 × .15 = .089, and
2. Ethnic ↔ FamBack × (FamBack → PrevAch → Grades + FamBack → PrevAch → Homework → Grades + FamBack → Homework → Grades) = .30 × (.42 × .44 + .42 × .22 × .15 + .17 × .15) = .067.

When added together, these equal .156, the implied correlation between Ethnicity and Grades. Another way to think about this is that the tracings listed under 1 are the total effects of Ethnicity on Grades, and those listed under 2 are the total effects of Family Background on Grades, times the correlation of Ethnicity and Grades.

3 I know properly it should be chi-squared, but, by convention, it's chi-square.

4 For example, type the χ^2 and *df* into two cells in Excel. Click on another cell, then Insert, Function. Click on CHIDIST and follow the directions to obtain the probability associated with χ^2 with the indicated *df*.

5 What do correlated disturbances mean? Focus on model C in Figure 13.14, which shows a correlated disturbance between d1 and d2. The disturbances represent all other influences on the corresponding variables other than those shown in the model. The correlation between d1 and d2 in this model suggests that the other influences (other than Ethnicity and Family Background) on Previous Achievement and Homework may be correlated. What this means, in turn, is that there may be other common causes of Previous Achievement and Homework not included in the model. Correlated disturbances can also be used to denote an agnostic causal relation; that is, we think that Previous Achievement and Homework are causally related but don't know the direction. As shown in Appendix C, one helpful way of thinking about partial correlations is that they represent the correlation between disturbances.

6 Model B and Models C and D resulted from the application of rule 2. Model E, with the paths between Homework and Ethnicity and Homework and Family Background reversed, also resulted from the application of this rule. Model F builds on Model E. Note that with model E Homework and Grades now have the same causes. We can therefore apply rule 3 to Model E and reverse the path from Grades to Homework. It may not be obvious, but Models E and F are nonrecursive models. Note that in Model F, for example, Homework affects Background, which affects Previous Achievement, which affects Homework, and so on.