Chapter 17

STRUCTURAL EQUATION MODELING[1]

Victoria Savalei, University of California, Los Angeles

Peter M. Bentler, University of California, Los Angeles

Introduction

Structural equation modeling (SEM) is a tool for analyzing multivariate data that has been long known in marketing to be especially appropriate for theory testing (e.g., Bagozzi, 1980). Structural equation models go beyond ordinary regression models to incorporate multiple independent and dependent variables as well as hypothetical latent constructs that clusters of observed variables might represent. They also provide a way to test the specified set of relationships among observed and latent variables as a whole, and allow theory testing even when experiments are not possible. As a result, these methods have become ubiquitous in all the social and behavioral sciences (e.g., MacCallum & Austin, 2000). A review of use of SEM in marketing research is provided by Baumgartner and Homburg (1996); see also Steenkamp and Baumgartner (2000).

In this chapter, we introduce the basic ideas of SEM using a dataset collected to test the theory of planned behavior (TPB; Ajzen, 1991). TPB is an extension of the earlier theory of reasoned action (Fishbein & Ajzen, 1975), and both attempt to address the often observed discrepancy between attitude and behavior. Theory of reasoned action proposed that attitudes influence behavior indirectly by influencing a person's intentions to act, and that subjective norms (for instance, what important others think is the appropriate behavior in a given situation) influence intentions as well. Theory of planned behavior expands theory of reasoned action by adding a third predictor of intentions: perceived behavioral control (PBC; see Figure 17.1 for a conceptual diagram). TPB is most applicable when

deliberative or planned behavior is of interest, and has been successfully applied to a wide variety of behaviors. Extensions of theory of planned behavior to include other predictive variables have also been proposed in marketing context (e.g., Bagozzi & Warshaw, 1990; Perugini & Bagozzi, 2001). In our example, we use the theory to understand predictors of dieting behavior.

The data were collected from a sample of 108 college students who had the goal of reducing or maintaining their current body weight.[2] Participants' attitudes towards dieting were examined using 11 semantic differential items, of which 6 were designed to tap the more cognitive aspect of the attitude (e.g., useless-useful) and 5 were designed to tap the more affective aspect of the attitude (e.g., unenjoyable-enjoyable). Subjective norms were measured by asking participants to list three most important persons to them and indicate how much each would approve or disapprove of their dieting. Perceived behavioral control (PBC) was measured by asking participants how much control they had over sticking to a diet, whether dieting was difficult or easy, and whether sticking to a diet for the next four weeks would be likely or unlikely. Intentions were measured by asking participants whether they plan to stick to a diet, whether they intend to stick to a diet, and finally whether they will expend effort to stick to a diet. A self-report measure of behavior was obtained a few weeks later by recontacting participants. Further details regarding the measures can be found in Perugini and Bagozzi (2001), who also test a more complex model.

The conceptual model for these data is illustrated in Figure 17.1. Here, each construct of interest is represented by a circle, and the influence of one variable on another is represented by an arrow. Assuming the relationships between the variables are linear, we can convert the diagram into two regression equations:

$$Int = \beta_1 Att + \beta_2 Norms + \beta_3 PBC + \varepsilon_I$$

$$Beh = \beta_4 Int + \varepsilon_B .$$

There are two equations because in the diagram, there are exactly two variables that have one-way

arrows aiming at them. Each β coefficient is a one-way arrow in the diagram. The residual ε's are not

shown.

[FIGURE 17.1 ABOUT HERE]

Thus, one way to estimate the coefficients β_1 through β_4 is by running two separate regressions

(see Chapter 13, Regression Models). There are several disadvantages to this method, however. First,

each construct appearing in the equation above, with the exception of behavior, is measured by several

different items in the dataset. To avoid fitting an extremely large number of regressions, we will have to

combine our items to create scales, thus losing any information about possible differential performance

of some measures of a given construct over others. Further, the newly created scales might or might not

be reliable, and while we can assess their reliability (using Cronbach's α, for example), this information

does not affect the computation of the regression coefficients, and we would interpret our final

regression equations as if they represented a good estimate of the relationships of interest. Finally, fitting

separate regression equations to different parts of our model does not provide any way to assess the

overall model fit. In other words, in addition to estimating what the values of the paths in our diagram

would be if the theory *were* true, we would like to test if the theory as presented in this diagram actually

is true. Some limited model testing occurs in a regression setting also; for example, when we look at the

scatter plot of our data to see whether it follows a curvilinear pattern, we informally evaluate whether

the assumed linear regression model actually holds, and this evaluation is quite independent of the

obtained estimates for slope and intercept.

Path Analysis

In contrast to the separate regressions approach, a statistical technique known as *path analysis*

(or *simultaneous equations*) can be used to obtain both the path values (estimated $\hat{\beta}$) for the model and

a test of the overall model fit. This technique is actually a special case of SEM, one that only involves observed variables, so we will discuss it in some detail. The goal of path analysis, and more generally of SEM, is to see how well our proposed model, which is a set of specified causal and noncausal relationships among variables, accounts for the observed relationships among these variables. The observed relationships are usually the covariances, summarized in the sample covariance matrix, which we will call S. If we could measure everyone in the population, we would obtain the population covariance matrix, Σ. Of course we cannot do that, but S serves as a good estimate of Σ, and this estimate gets better as the sample grows larger. The most important idea in SEM is that under the proposed model, the population covariance matrix Σ has a certain structure; that is, some of its elements are functions of other elements or other parameters in the model (such as regression coefficients). If we estimate these more basic parameters from the data, we can compute an estimate of the population covariance matrix (call it $\hat{\Sigma}$) that is based on the assumed model as well as the data. When the model is true, S and $\hat{\Sigma}$ are estimates of the same thing, namely Σ. When the model is false, they are not. Thus, we can evaluate model fit by comparing S and $\hat{\Sigma}$ as estimated from our sample.

Before we can illustrate this idea on our example, we need a more detailed diagram for our model. In Figure 17.2, we follow the path analysis convention of using boxes to represent observed variables (circles are reserved for latent constructs), one-way arrows to represent regression coefficients, and curved double-headed arrows to represent covariances among independent variables. Here, we chose to allow the independent variables on the left – the variables without any one-way arrows aiming at them -- to correlate, which is a common assumption. However, a model without such correlations or only with some of them present can just as easily be specified and tested. Random errors from the regression equations we stated earlier are now also part of the diagram. For example, the equation $Beh = \beta_4 Int + \varepsilon_B$ states that behavior is influenced both by intentions and by random error (which can

also be thought of as all other unspecified influences). In the diagram, this equation is represented by two arrows pointing to behavior, one from intentions and one seemingly from nowhere. Technically, the random errors themselves should be represented as variables in the diagram, but it is common to omit them. Paths that are *not* in the diagram are especially informative of the model being tested. For example, the random error from predicting behavior is uncorrelated with any of the other independent variables, because there are no curved arrows starting from it. Also, there is no direct influence of attitudes on behavior; the relationship between attitudes and behavior is fully mediated by intent. When we test this model, we are testing the plausibility of all these assumptions simultaneously.

[FIGURE 17.2 ABOUT HERE]

For the five variables in our example (attitudes, norms, PBC, intentions, and behavior) the sample covariance matrix looks like this:

$$S = \begin{pmatrix} 1.10 & & & & \\ 0.72 & 2.98 & & & \\ 0.69 & 0.73 & 4.48 & & \\ 0.91 & 0.93 & 0.52 & 2.60 & \\ 0.38 & 0.46 & 0.10 & 0.59 & 0.93 \end{pmatrix}$$

Here, each variable is created by averaging all the items that measure it. This matrix is symmetric, with variances of each scale on the diagonal and their covariances below the diagonal (because of the symmetry, we do not display the elements above the diagonal). There are a total of 15 distinct elements: 5 variances and 10 covariances. In SEM, these are the "data points": the unique pieces of information we have to evaluate whether the population covariance matrix has a certain structure. There will always be $p(p+1)/2$ unique elements in a covariance matrix of p variables. We can represent the population covariance matrix as:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & & & \\ \sigma_{12} & \sigma_2^2 & & & \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & & \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 & \\ \sigma_{15} & \sigma_{25} & \sigma_{35} & \sigma_{45} & \sigma_5^2 \end{pmatrix}$$

For example, σ_1^2 is the variance of the attitudes scale in the population, σ_{12} is the covariance of the attitudes scale with the norms scale in the population, and so on. As with the sample covariance matrix, the population covariance matrix has 15 distinct elements.

If our proposed model holds in the population, however, we do not need quite so many elements to describe Σ. For example, consider σ_{14}. Using the equation for intentions implied by the diagram, we can write:

$$\sigma_{14} = \text{cov}(Att, Int) = \text{cov}(Att, \beta_1 Att + \beta_2 Norms + \beta_3 PBC + \varepsilon_1)$$

Using the rules of covariance algebra,[3] we can rewrite this as $\beta_1 \sigma_1^2 + \beta_2 \sigma_{12} + \beta_3 \sigma_{13}$, because the covariance of any variable with itself is its variance, and because random error is not correlated with anything. Thus, we have shown the *covariance structure* of σ_{14} to be a function of other elements of Σ and of the regression coefficients. If we did the algebra for every element of Σ, we would find that the covariance structure of the entire matrix is a function of 12 things: the four regression coefficients, the five variances of the independent variables (counting variances of the errors, which we denote by ψ_4 and ψ_5), and the three covariances among the independent variables. These are called *model parameters*, and are usually summarized in one vector θ. That is,

$$\theta = (\beta_1, \beta_2, \beta_3, \beta_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \psi_4, \psi_5, \sigma_{12}, \sigma_{13}, \sigma_{23}).$$

Several methods exist by which we can estimate θ from our data; most of them are analogous to least-squares methods in regression, whereby some function of the residuals between the observed

covariances (elements of S) and the model-expected covariances (elements of $\hat{\Sigma}$) is minimized. This function is called the *fitting function*, indicated by F or $F(S, \hat{\Sigma})$ to stress its reliance on the comparison of two matrices. We will discuss this function in more detail in the next section; for now, assume we were able to obtain an estimate of θ somehow and thus to compute $\hat{\Sigma}$. In our example, using a popular method of maximum likelihood, we get:

$$\hat{\Sigma} = \begin{pmatrix} 1.10 & & & & \\ 0.72 & 2.98 & & & \\ 0.69 & 0.73 & 4.48 & & \\ 0.91 & 0.93 & 0.52 & 2.60 & \\ 0.20 & 0.21 & 0.12 & 0.59 & 0.93 \end{pmatrix}$$

The matrix of *residuals* is given by:

$$S - \hat{\Sigma} = \begin{pmatrix} 0.00 & & & & \\ 0.00 & 0.00 & & & \\ 0.00 & 0.00 & 0.00 & & \\ 0.00 & 0.00 & 0.00 & 0.00 & \\ 0.17 & 0.25 & -0.02 & 0.00 & 0.00 \end{pmatrix}$$

We see that the covariances between attitudes and behavior and norms and behavior are not very well explained by the model. Converting these differences to correlations for easier interpretation, we obtain that the two largest *standardized residuals* are 0.17 and 0.15. Whether or not these are too big to declare the proposed model implausible depends on our sample size. If the sample size is large, we can expect that S is a pretty precise estimate of Σ and large residuals are due to the model being wrong and not to sampling fluctuations. In our example, N=108, which is actually not very large, given that we have estimated 12 parameters. So perhaps residuals of this size are likely to occur even if the model holds.

To formally answer this question, we compute the following test statistic: $T = (N-1)F(S, \hat{\Sigma})$, which, under certain assumptions, approximately follows a chi-square distribution. The hypothesis we

are testing is the covariance structure hypothesis that $\Sigma = \Sigma(\theta)$, that is, the covariance matrix is a

function of just 12 elements (those in θ). The degrees of freedom for this test is the number of distinct

elements in the covariance matrix minus the number of model parameters. In our example,

$df = 15 - 12 = 3$, $T = 7.70$, and the corresponding p-value is 0.053, using a chi-square distribution with

three degrees of freedom. We can interpret this as follows: under the assumption that our model is true,

the probability of observing the residuals as large as or larger than ours is about 0.053. A curious thing

about SEM (and all model testing in general) is that we actually want to retain our hypothesis of a

certain model structure, and would be happy with a large and not a small p-value. In this example, the

data seem to only marginally support the model. We do not present the actual obtained parameter

estimates (that is, the estimated values of θ).

From the definition of degrees of freedom, we see that the one kind of model we cannot test is

that which imposes no structure on the covariance matrix because in this case we would have zero

degrees of freedom. This model is called *saturated*. In our example, we can obtain the saturated model

by adding paths from attitudes, norms, and PBC to behavior, so that in the resulting diagram every

observed variable is related to every other observed variable. If we estimate such a model, we would

obtain parameters identical to those from ordinary multiple regressions, and no test of model fit. In fact,

if we take another look at the residual matrix from our example, we see that the relationships among the

first four variables are reproduced perfectly; this is because the part of the model that relates these four

variables is in fact saturated. Thus the only "testable" part concerns the relationship between behavior

and the three predictors of intention, and this is the part that is only marginally supported by the data. At

the other extreme from the saturated model we have the *independence model*, which assumes that all the

variables are uncorrelated with each other. The model-implied covariance matrix in this case is a

diagonal matrix (that is, all the off-diagonal elements are zero), and we would only have to estimate the

five diagonal elements (the variances of the variables). In our example, this leads to 15-5=10 degrees of freedom. Thus, this model is testable, but not very interesting, although we will see some uses for it later.

First Look at SEM

Path analysis clearly has advantages over performing a series of multiple regressions; namely, it provides a test of the overall model fit. But it still possesses some of the same disadvantages, the biggest one being that it does not take into account the reliability of observed variables and treats them as perfect substitutes for the constructs they represent. A full-blown structural equation model solves this problem by representing each construct as a latent variable (also called a factor). Alternative viewpoints about latent variables are reviewed by Bollen (2002). A latent variable explains the relations among observed variables (indicators) that measure the construct. This prediction does not have to be perfect, so that the reliability of each indicator as a measure of the latent construct can be estimated. Figure 17.3 gives the diagram for a sample structural equation model that could be tested with our data. It has four latent constructs, each measured by three indicators (we actually have 12 indicators of attitudes, but we only use three for the purposes of this illustration). The relationships among these constructs constitute the *structural* part of the model. The *measurement* part of the model consists of the relationships between the latent variables and their indicators, and the values of these paths are referred to as *loadings* (see Chapter 18, Cluster Analysis and Factor Analysis). Note that each indicator has an error term associated with it, which allows for imperfect measurement. These error terms are not correlated because we assume that the different measures of the same construct are only related because of their dependence on the underlying construct (this assumption can sometimes be relaxed in practice). Finally, one of the loadings for each latent variable is fixed to 1. This is because latent variables, by virtue of being entirely imaginary, do not have set scales, and thus need to be assigned some arbitrary units.

Picking a good indicator of the latent construct is one option; another common solution is to set the variance of the latent variable to 1. A somewhat unusual feature of our model is that behavior has remained an observed variable, because only one measure was available for it (namely, self-report). This of course does not mean that self-report is a perfectly reliable indicator of the actual behavior, but only that alternative measures were not obtained in this study, perhaps because of difficulty or cost.

All of the concepts defined for path analysis generalize completely to structural equation models. For instance, the degrees of freedom for a model are again the number of unique elements in the covariance matrix minus the number of estimated parameters. For the model of Figure 17.3, we have 13 observed variables, and thus $(13*14)/2 = 91$ unique elements in the sample covariance matrix. From the diagram we can deduce that the estimated parameters will be: variances of latent variables (3), error variances (11), covariances among the latent variables (3), regression coefficients or loadings (12), resulting in a total of 29 model parameters. Thus, $df = 91 – 29 = 62$. As in path analysis, this model implies a certain covariance structure, and we can test whether our sample covariance matrix S roughly follows this structure, given our best estimates of the model parameters, obtained by minimizing some fitting function. Finally, our model implies a set of fourteen regression equations (one for each of the observed variables, and one for the latent construct of intentions), which we will not state here.

It is worth noting that models such as Figure 17.3 can be looked at in two different ways. The traditional way, developed in the LISREL program (Jöreskog & Sörbom, 1994), is to consider separate sets of equations for the measurement model and for the structural model. This approach requires the use of eight Greek-labeled matrix equations to specify a model. An alternative way (Bentler & Weeks, 1980), used in the EQS (Bentler, 2005) program, is to provide an equation for every dependent variable and covariances for independent variables as illustrated below.

[FIGURE 17.3 ABOUT HERE]

Having given a brief conceptual introduction to SEM using the simpler idea of path analysis, we now discuss the process of structural equation modeling in more detail, with our TPB example as an illustration.

The Modeling Process

The process of modeling involves four general stages: specification, estimation, evaluation, and modification. In the specification stage, we develop the model we want to test and convert this information into a format that a computer program can understand. In the estimation stage, we choose a fitting function and obtain parameter estimates for our model. In the evaluation stage, we interpret the test of model fit and other indices of fit. In the modification stage, we modify the original model in accordance with the information obtained in the previous stage as well as theory. We now discuss each stage in more detail and provide an illustration using our TPB example. We use the SEM computer program EQS 6.1 to do the computations.

Model Specification

As we saw earlier, not every model we come up with can be tested (in particular, saturated models have zero degrees of freedom). Furthermore, some models we might accidentally come up with cannot even be estimated, let alone tested—that is, when we try to use a least-squares or some other criterion to get parameter estimates, we obtain no unique solution. This is the problem of identification, and should be addressed during model specification stage. A model is *identified* if we are able to obtain a unique solution for every parameter. Unfortunately, this condition is hard to verify for an arbitrary model, but we can be fairly sure it is met by following a few simple rules. First, an identified model must have nonnegative degrees of freedom; that is, number of estimated parameters should be less than or equal to the number of data points obtained from the sample covariance matrix. Second, every latent variable in the model needs to be assigned a scale; this is usually accomplished by fixing one of its

loadings to one. Third, the latent variables need to relate to a few other things to allow their

identification; after all, these are imaginary constructs and we need to get at them somehow. A latent

construct with three indicators will be identified; two indicators can work if there is also a nonzero

correlation with another construct in the model, or if additional constraints are imposed on the loadings

of the indicators.[4] More complex identification rules can be found in Bollen (1989). However, once the

necessary conditions for identification stated above have been checked, the easiest way to see whether

the model is identified is to run it through an SEM program and look for any error messages.

The model in Figure 17.3 appears to meet the identification conditions and could easily be

adapted to our data if we added more indicators to the attitudes factor (as a reminder, our dieting dataset

has 21 variables: 11 attitude items, 3 norms items, 3 PBC items, 3 intention items, and 1 measure of

behavior). However we make a few other changes. First, we combine the 11 attitude items into 6

composites, where each composite is the average of two items, except for the very last item which is left

intact. This procedure is known as *item parceling*, and its primary purpose is to reduce the complexity of

the model.[5] Although in theory the more indicators a latent variable has the better, in practice large

number of observed variables can make the model too difficult to estimate successfully. If the

researcher's interest is in the structural model, item parceling can reduce the problem of model

complexity by simplifying the measurement model while keeping the structural model intact. In

addition, item parcels are likely to have smoother distributions and higher reliabilities than the original

items. Excellent summaries of pros and cons of item parceling are given by Bandalos and Finney (2001)

and by Little et. al. (2002).

Second, we model two separate attitude components: a cognitive (called "evaluative" in Perugini

& Bagozzi, 2001) and an affective component, each measured by three indicators. This partition is

theoretically appropriate because the first six attitude items were written to measure cognitive evaluation

of dieting and the last five items were written to measure affective reactions to it (see Kim, Lim, & Bhargava, 1998, and Bodur, Brinberg, & Coupey, 2000, for support of such a view of attitudes, and Fishbein & Middlestadt, 1995, for an opposing view). An exploratory factor analysis of all 11 items provided empirical support for this partition (see Chapter 18). Because in practice the cognitive and the affective component of an attitude are often highly correlated (Eagly, Mladinic, & Otto, 1994; Trafimow & Sheeran, 1998), we also introduce a *second-order* factor that represents the overall attitude and predicts both the cognitive and the affective components. Our initial model allows attitudes to influence intentions only via this second order attitudes factor (see Figure 17.4).

[FIGURE 17.4 ABOUT HERE]

Finally, we also replace the latent variable PBC with just one of its indicators, turning it into an observed variable. This change is based on the examination of the correlation matrix for the data, given in Table 17.1. In our theoretical model, PBC should predict intentions to diet. The correlations between the first PBC item and the three intentions items are 0.36, 0.40, and 0.24, suggesting that this item functions as intended. However, the correlations between the other two PBC items and the three intention items are 0.07, 0.17, 0.01, -0.01, 0.02, and -0.12. Most of these are very small, and the negative values are particularly troublesome. Ideally, the researcher should carefully examine the wording of these items and build some hypotheses as to why they do not function as expected. However, for the purposes of our illustration, we simply use the first item as a proxy for the construct of perceived behavioral control. The diagram of the final model with these three changes incorporated is given in Figure 17.4.

[TABLE 17.1 ABOUT HERE]

We now specify and run this model in EQS 6.1. Due to space limitations, we cannot give a thorough introduction to EQS (or any other computer program) and will provide its input and output

primarily to illustrate the modeling process. See Byrne (1994) for an introduction to EQS. Other

software packages are listed at the end of this chapter. Table 17.2 gives sample EQS syntax for the

model in Figure 17.4. The code is broken into several sections. In the Specifications section, we provide

details such as the name of the data file, the number of variables and cases, and the estimation method.

By default, EQS uses V's to label all observed variables, and F's to label all latent variables. Different

names can be provided in the Labels section. This is a good idea because the new names will be used in

the output, aiding in its interpretation. Model specification in EQS involves providing equations for each

dependent variable and statements about variances and covariances of the independent variables.[6] The

Equations section contains 16 equations, one for every dependent variable in the diagram. Some paths

are fixed to one for identification purposes; asterisks indicate paths that are estimated. E's represent the

errors associated with the prediction of observed variables, and D's (disturbances) represent errors

associated with the prediction of latent variables. The Variances section contains specifications of the

variances of the independent variables. In our example, the variances of all independent variables,

whether latent or observed, are freely estimated. Note that E's and D's are also independent variables.

The Covariances section lists the covariances among the independent variables to be estimated. The

three double-headed arrows in Figure 17.4 have been converted into three lines of code; the rest of the

covariances are fixed by the program to zero. In particular, note that E's or D's are assumed not to

correlate with anything, which is a standard assumption, as the errors are conceptualized as being

entirely random.

[TABLE 17.2 ABOUT HERE]

Model Estimation

In the estimation stage, we choose a fitting function and minimize it to obtain parameter

estimates. This is an *iterative* process: we first plug in the initial values for all the parameters and

evaluate the function, then we modify the parameter estimates in an attempt to make the function

smaller, we then reevaluate the function, and so on, until the value of the function no longer changes by

much from one iteration to the next (this is called *convergence*). Because this process is impossible to

carry out by hand, the choices available to the researcher during estimation largely depend on the

software used. Estimation methods available in EQS include ML, LS, GLS, and AGLS. Of these,

maximum likelihood (ML) is by far the most popular and is the method we recommend. However, the

equation for the ML fitting function is also the least intuitive. Thus, we discuss other fitting functions

first.

Recall that a fitting function is a summary measure of the size of the residuals in the model. The

simplest such function is the sum of squared residuals, or the LS (least-squares) fitting function. If s_i

and $\hat{\sigma}_i$ are all the unique elements of S and $\hat{\Sigma}$ respectively, the LS function looks like this:

$F_{LS} = \sum_i (s_i - \hat{\sigma}_i)^2$. The parallel equation in regression is the least-squares criterion: $\sum_i (y_i - \hat{y}_i)^2$, which

minimizes the sum of squared residuals between the observed and predicted values of y . In the

regression setting, this criterion is only optimal if the assumption of homoscedasticity is satisfied. When

this assumption is violated, weighted least-squares (WLS) regression can be used instead, which

minimizes a weighted sum of squares, with the weights reflecting the different variances of individual

elements. In SEM, the assumption of homogeneity is never plausible, because in the place of y we have

the very different elements of the sample covariance matrix, whose variances have no reason to be the

same (these "variances of the variances," or fourth-order moments, are related to each variable's

kurtosis). Moreover, while in regression we assume that the observations are independent of each other,

in SEM the elements of the sample covariance matrix are not in fact independent, and additional weights

related to their covariances also need to be estimated. Thus, in SEM, LS estimation is rarely the optimal choice.

Most fitting functions, such as GLS and AGLS, are loosely analogous to the weighted least-squares procedures in regression, and in fact are often called WLS estimators in the literature (when this term is used, it pays to find out which particular method is being referred to).[7] These "generalized" least squares methods differ in the assumptions the researcher must make about the data and in the choice of weights. GLS is appropriate when the variables have no excess kurtosis, so that the weights are greatly simplified. This estimator is appropriate when the data are normally distributed, for example. AGLS ("arbitrary distribution" GLS) does not require any assumptions and estimates all the weights from the data before using them in a fitting function. Because estimating these weights accurately requires large samples, this method almost never works well unless the sample size is very large (perhaps a thousand or more) or the model is very simple. AGLS is also known as ADF, or "asymptotically distribution free," in the literature.

The ML fitting function has a different and more appealing rationale, but one that requires the assumption that the joint distribution of the data is multivariate normal. If such an assumption is made (we will discuss how to evaluate its plausibility in the next section), the ML parameter estimates maximize the likelihood of observed data under the estimated model. The function F_{ML} is unenlightening,[8] but it actually does something very similar to minimizing a weighted sum of squared residuals, where the weights are constructed using the normality assumption and updated in each iteration. As another point of comfort, ML estimates are usually the most precise (minimum variance) estimates available.

Despite the restrictive normality assumption, the ML parameter estimates are actually fairly robust to the violation of this assumption, and ML is the preferred method of estimation even if this

assumption is violated. However, the standard errors for parameter estimates as well as the model chi-square are affected by nonnormality. But as we will see, standard errors and the chi-square can be adjusted when the data are nonnormal, and these adjustments coupled with the ML parameter estimates turn out to work better in practice than many other estimation methods that require fewer assumptions. In EQS, we specify METHOD=ML to obtain ML parameter estimates, as is done in the syntax in Table 17.2. We defer the discussion of the EQS output for parameter estimates until we have discussed model evaluation and found a well fitting model.

Model Evaluation

There are two components to model fit: statistical fit and practical fit. Statistical fit is evaluated via a formal test of the hypothesis $\Sigma = \Sigma(\theta)$, whereby we compute a test statistic and the associated p-value. Practical fit is evaluated by examining various *indices of fit*, which attempt to summarize the degree of misfit in the model. For example, the average standardized residual is one fit index that can help decide whether the model provides a good enough approximation to the data. Statistical fit is analogous to a p-value in an ANOVA setting, and fit indices are analogous to effect size measures. The debate about the relative virtues of statistical significance versus practical significance prevails in the SEM setting as well, with one big difference: in the ANOVA setting, this debate is usually over the relative importance of statistically significant findings with trivial effect size, whereas in SEM, the debate is over acceptance of models with trivial nonzero residuals and a statistically significant chi-square.

The hypothesis $\Sigma = \Sigma(\theta)$ is formally evaluated using the statistic $T = (N-1)F(S,\hat{\Sigma})$, which, if the assumptions of the estimation method are met, has an approximate chi-square distribution with $p(p+1)/2 - q$ degrees of freedom. Here, N is sample size, $F(S,\hat{\Sigma})$ is the minimized value of the fitting function, p is the number of variables, and q is the number of estimated parameters. As we have

already mentioned, an unusual aspect of SEM testing is that the null hypothesis $\Sigma = \Sigma(\theta)$ is actually the

hypothesis we want to *retain*. We therefore want T to be small. Further, the familiar $\alpha = 0.05$ criterion

is also used here to retain or reject models; for example, a p-value of 0.06 is interpreted as evidence in

support of the model, despite the fact that its meaning remains the same: if we assume that the model is

true, the probability of observing residuals as large as or larger than ours is only 6%! Thus, what is a

stringent criterion in ANOVA appears like a rather sloppy one in SEM. However, it is not so easy to

obtain a model that passes the chi-square test even using this liberal criterion, and in particular it gets

more difficult as the sample size gets large. This is because the sample size multiplier $(N-1)$ enters the

equation for T. For model residuals of the same size, the larger the sample, the larger the test statistic.

In other words, statistical power works against us in the cases when we want to "prove" the null

hypothesis. This peculiarity about the model chi-square test is why alternative fit indices are often

considered along with it.

There are many different kinds of fit indices; we discuss the most popular ones, and then

recommend two in particular. Perhaps the most intuitive measure of practical fit is standardized root

mean-square residual (SRMR). This index is equal to the square-root of the average squared element of

the residual correlation matrix. A popular cut-off value for this index is 0.05 or less.[9] However, if the

residual matrix has many elements, this index can mask big standardized residuals and create the false

impression that $\hat{\Sigma}$ is a good approximation to S. For this reason, examining raw residuals is also useful.

As will be seen shortly, part of the standard EQS output is a list of twenty largest standardized residuals

in decreasing order.

Some fit indices are based on the idea of estimating the "proportion of variance" in the observed

data that is explained by the model. The simplest one of these, the Goodness of Fit index (GFI), is equal

to one minus the ratio of the residual weighted sum of squares (using elements of $S - \hat{\Sigma}$) over the total

weighted sum of squares (using elements of S), where the weights are as in the fit function. This index is directly analogous to R^2 in ordinary regression. Another index, AGFI (Adjusted GFI), is analogous to the adjusted R^2. Both GFI and AGFI take on values between 0 and 1, with values less than 0.90 often considered unacceptable. A drawback of both these indices is that they tend to produce somewhat higher values as the sample size gets larger, despite their goal of providing an alternate measure of fit that is independent of N.

Other fit indices, called incremental fit indices, assess practical fit by considering the improvement in fit over a baseline model, usually the independence model (where $\hat{\Sigma}$ is a diagonal matrix). Denoting the model chi-square by χ^2_M and the chi-square associated with the independence (null) model by χ^2_N, these indices measure the relative change in the chi-square between the independence model and the tested model. The simplest index that does this is the Normed Fit Index (NFI), which is simply equal to the percentage change in the chi-square: $NFI = \dfrac{\chi^2_N - \chi^2_M}{\chi^2_N}$. Even though the sample multiplier N does not explicitly enter the equation for NFI, this index, too, tends to be too small for models based on few observations. Bollen's Incremental Fit Index (IFI), defined as

$$IFI = \dfrac{\chi^2_N - \chi^2_M}{\chi^2_N - df_M},$$ tries to correct for this dependence on sample size. Yet another index, called NNFI (Nonnormed Fit Index, also known as Tucker-Lewis Index or TLI) is a modification of NFI that rewards parsimonious models, but it can take on values higher than 1, making it somewhat difficult to interpret.

Many other fit indices have been proposed. A possible reason for such an abundance of fit indices is that no one index has been able to meet all the criteria researchers want it to meet: to have a finite range (e.g., 0 to 1), to reward models that are "far" from the independence model, to reward parsimonious models (models with many degrees of freedom), to be independent of sample size (in

contrast to the chi-square), and to have a clear and well-established cut-off value (such as 0.90 or so). For this reason, multiple fit indices should be examined and reported when evaluating practical fit of a model.

We now introduce two more fit indices that have become widely accepted in the field. The first one is called the Comparative Fit Index (CFI) and is given by $CFI = 1 - \dfrac{\chi_M^2 - df_M}{\chi_N^2 - df_N}$. If $\chi_M^2 < df_M$, the value of CFI is constrained to 1 (this is not an interesting case, however, because such a model will also pass the chi-square test). When the model is correct, the expected value of the test statistic is its degrees of freedom, so that $\chi_M^2 - df_M$ is close to zero, and CFI is close to 1. When the model is not correct, the expected value of the test statistic is approximately the degrees of freedom plus the *noncentrality*, a parameter of the chi-square distribution usually indicated by λ . The CFI index can be thought of as a measure of relative noncentrality between the tested model and the independence model, because we can rewrite it as $CFI = 1 - \dfrac{\hat{\lambda}_M}{\hat{\lambda}_N}$, where $\hat{\lambda}$ represents an estimate of the noncentrality for each model.

Another index based on the noncentrality is the RMSEA, or root-mean squared error of approximation, given by $RMSEA = \sqrt{\dfrac{\hat{\lambda}_M}{(N-1)df_M}}$. It measures the average amount of misfit in the model per degree of freedom. In contrast to most other indices, smaller values indicate better fit. In practice, RMSEA and CFI are often used together to judge model fit; a popular criterion is to accept models that have CFI>0.90 and RMSEA<0.05. Hu and Bentler (1999) recommend a more stringent CFI>.95, and a less stringent RMSEA<.06. It should be noted that fit indices are like R^2 values: there is no magic value at which fit is good enough. Because they are not tests of correct model specification, they should be used with caution. For a serious critique of reliance on fit indexes to judge model fit, see Marsh, Hau, and Wen (2004). We now illustrate evaluating statistical and practical fit using our dieting example.

Table 17.3 gives selected EQS output for the syntax in Table 17.2. We first examine the residuals. From the standardized residual matrix, we see that about half of the residuals are positive and half are negative (which is a good indicator of randomness) and most are pretty small. The variable that stands out, however, is V1, the first indicator of the cognition factor. Its correlations with the indicators of intention (V14-V16) and with behavior (V13) are not very well explained by the model. In fact, V1 is involved in the top four largest residuals. To see whether these large residuals can be due to sampling fluctuations, we turn to the chi-square test, which is reported in the Goodness of Fit Summary. The chi-square value is 107.6, with 70 degrees of freedom and the associated probability of .003. This means that, even with our small sample size (N=108), the residuals as large as the ones we are observing are not likely under the proposed model. The fit indices paint a mixed picture. In particular, CFI looks very good (.954), but RMSEA is a little large (.071), and a few other fit indices are below 0.90. The standardized root mean-square residual, or SRMR, is .075. Taken together, the residuals, the chi-square, and the fit indices lead us to conclude that our model in its current form does not fit the data, and therefore we do not report the parameter estimates. Instead, we proceed to modify our model so that it better accounts for the observed covariances.

[TABLE 17.3 ABOUT HERE]

Model Modification

When the original proposed model does not fit the data, we can consider what modifications, if any, will help improve its fit. The exploratory nature of such an analysis should be acknowledged, as we are "peeking at the data" to find a well-fitting model. But there is nothing wrong with trying to find a set of relationships that explains the observed covariances. After all, data can be expensive to obtain, and throwing it out without fully discovering what it has to "say" is not the wisest thing to do. Modifications

to the model can be suggested by the residuals obtained in the original run as well as by special statistics called *modification indices*. These indices point specifically to paths whose addition to the model would result in the biggest improvement in the overall chi-square value. Of course, these modifications need also to make sense theoretically if we are to interpret the resulting model. It should also be noted that while many researchers modify the model until it passes the chi-square test, the resulting p-value is not fully meaningful because it was obtained by "peeking at the data." The actual probability of observing the resulting residuals is probably lower than the p-value for the modified model. In a perfect world, the modified model would be cross-validated on a new sample, especially if the original sample was small.

The kind of modifications we consider here are adding paths or dropping paths. More global modifications are of course possible (such as reworking the theory entirely), but will probably not be driven by statistical information. Further, because dropping paths cannot improve fit, when we start with an ill-fitting model we only consider adding paths. We will return to the idea of dropping paths once we have a model that fits the data. As we have already seen, the biggest unexplained correlations under our original model are between V1 and the three measures of intentions. Therefore, we choose to add a path from V1 to the latent variable F5, which represents the construct of intentions. There are two ways to do so. One is to add a path directly from V1 to F5; the other is to add a path from E1 to F5, where E1 is the residual of V1 or the portion of V1 independent of F1. We choose the second approach, essentially adding a predictor orthogonal to the set of variables already influencing F5. A direct path from E1 to F5 is not inconsistent with the theory of planned behavior; it simply means that the two items that compose the first indicator of the cognition factor predict intentions over and above the influence captured by the cognition factor. From Table 17.1, we see that indeed the correlations between V1 and the intention variables are all in the .6 range, whereas the correlations between V2 and V3 and the intention variables are in the .4 range. The additional path captures this differential influence. Further, since intentions

predict behavior, the effect of E1 on behavior is mediated by intentions (see e.g., MacKinnon et al., 2002, for a discussion of mediation). The residual correlation between V1 and behavior was also large, but adding a path from V1 directly to V13 (behavior) would modify the theory of planned behavior in a more substantial way, because it would allow attitudes to directly influence behavior. Thus, we choose not to add this path for now. It is also good practice to modify the model one step at a time.

Although we already have a pretty good idea how we want to modify our model, a more formal way to determine the next step in model modification is to examine the Lagrange Multiplier (LM) tests (Chou & Bentler, 1990), which are one type of modification indices (Sörbom, 1989). These are given in the very last part of the output presented in Table 17.3. The values in the table represent the predicted drop in the model chi-square that would result from adding the specified path to the model. The first listed parameter, (F5,V1), represents a causal path from V1 to F5, which is exactly the path we were planning on including into our model based on the examination of the residuals. We see that if we add this path, the model chi-square is expected to drop by about 27 units (given that it was about 107 before, this may be enough to reach statistical "insignificance"), and the included path will take on the standardized value of about .40, which is pretty large.

When considering LM tests, we have to be careful to ignore paths that are theoretically meaningless. As an illustration, consider the second largest LM test, associated with the path (V1, F5). This LM test suggests that the factor of intentions should influence one of the indicators of attitudes, which goes against the proposed flow of causality. Even if the LM test for this path happened to be the largest, we would not have chosen to include this path. Examining other entries in the table also suggests some paths that we would not have easily thought of based on the examination of residuals only. For instance, a path from V4 to F5 has the third largest predicted chi-square drop (of about 7.8). It should be mentioned, however, the LM tests reported in this table are *univariate*: they estimate changes to the chi-

square without taking into account previously proposed changes. Thus, once we add a path from V1 to F5, the LM test value for other paths might change. This is why it is best to conduct model modification one step at a time.[10]

To run the modified model in EQS, we only need to change one equation in the syntax in Table 17.2. We replace the equation for F5 with the following:

F5 = *F3 +*V7+*F4+*E1+D5. Selected EQS output for this run is given in Table 17.4. From the table of largest standardized residuals, we see that the largest misfit now occurs in accounting for the correlation between behavior (V13) and indicators of various factors predicting intentions (such as V10 and V2). These residual correlations are in the 0.17-0.20 range, which is still not small (although smaller than those observed in the original model, which were about 0.3), leading us to wonder whether the factors affecting intentions do not also influence behavior directly. However, the chi-square for the modified model is now 76.01 with 69 degrees of freedom and the associated p-value is 0.25, suggesting that even residuals as large as 0.2 are entirely within sampling fluctuations given our small sample. Of course, this observed p-value is conditional on our having peeked at the data, and the true probability might be lower. There is little we can do about this problem, and only cross-validating this model on a new sample can answer the question of whether the added path is truly there and whether direct paths from factors affecting intentions to behavior are necessary. The modified model also fits well by other criteria; in particular, CFI=.99 and RMSEA=0.03. Finally, we also note that the independence model does not fit the data, $\chi^2(91) = 912.7$. While this observation seems trivial, with small sample sizes such as ours, it can happen that we lack power to reject even the most restrictive model, rendering other model testing meaningless.

[TABLE 17.4 ABOUT HERE]

Parameter Testing

Once we have a model that fits the data well, we can proceed to interpret parameter estimates and to test for their statistical significance. We discuss several such tests, in particular z-tests, Wald tests, and chi-square difference tests. We do so in the context of interpreting the output for our model.

In Table 17.4, parameter estimates are given separately for measurement equations (those where the DV is an observed variable) and for construct equations (those where the DV is a latent variable). In each equation, below the parameter value is its standard error, followed by a kind of a z-statistic, which is the ratio of the parameter estimate to the standard error. If this z-statistic is significant at the 0.05 level, a @ symbol is printed next to it. Some factor loadings do not have standard errors, because they were fixed a priori to one and not estimated. Coefficients in front of the error terms were also set to one as is done in ordinary regression (recall the equation $y = \alpha + \beta x + \varepsilon$, with an implicit 1 in front of the error term). Variances and covariance estimates for the independent variables are given in a separate section.

We first observe that each estimated factor loading is statistically significant, and the measurement part of our model seems to function as it should. Behavior (V13) is significantly predicted by the factor of intentions (F5), which is consistent with the theory of planned behavior. However, the construct of intentions itself is only predicted by two variables: F3, the overall attitudes factor, and E1, the residual term associated with V1, which was the path we added during model modification. Neither norms (F4) nor perceived behavioral control (V7) have a significant effect on intentions. A bigger sample size might render these effects reliable. We next examine the variances of the independent variables. Usually variances are uninteresting, but the variance of D1, the residual term associated with the prediction of the cognition factor (F1) from the overall attitudes factor (F3), is not significantly different from zero. This could mean that the overall attitudes factor is somewhat indistinguishable from the cognition factor, and the second-order factor structure might be superfluous here. Finally, the

covariance between PBC and the Norms factor is also not significantly different from zero, but with no a priori hypotheses about correlations among predictors of intentions, we are not especially interested in this finding.

The most useful part of the output in Table 17.4 is the standardized solution, where all variables have been transformed to unit variance. This means that paths are now standardized partial regression coefficients (often called beta weights in regression), and covariances are correlations. The standardized paths are given in Figure 17.5. Because each observed variable is predicted by only one factor, factor loadings are now on a metric we can understand: they are just the correlations between the variables and the factors. We see that all of the loadings look pretty good, ranging from 0.53 to 0.93. The R^2 values given next to standardized equations for the measured variables represent the proportion of variance in each variable that is explained by the factor it loads on, and, for those measured variables that are conceptualized as indicators, can be thought of as reliabilities. We see that about 18% of the variance in behavior is explained by the factor of intentions. This is not very high, but fairly typical of this area of research. In contrast, almost 59% of the variance in the factor of intentions is explained by its predictors. The strongest predictors of this factor are the two attitudes variables, F3 and E1, and these paths are fairly large (0.47 and 0.46). The standardized paths for V7 (PBC) and F4 (Norms) are 0.107 and 0.136, but, as we know from the unstandardized output, these values are not reliably different from zero. Finally, the nonsignificant correlation between norms and PBC is actually .228, and a bigger sample size might detect this relationship.

[FIGURE 17.5 ABOUT HERE]

When a nonsignificant path exists in an otherwise well-fitting model, we may ask whether the model would fit the data equally well or about as well if we were to omit this path entirely.[11] We can formally answer this question by means of a chi-square difference test. This test requires that in addition

to the model we already have, we also estimate the more restricted model by excluding the path (or paths) of interest. We choose the nonsignificant path from norms to intentions to illustrate this idea. When we estimate the model with this path set to zero, we obtain a chi-square value of 77.94 with 70 degrees of freedom. Our original chi-square value was 76.01 with 69 degrees of freedom, and we obtain the chi-square difference test by simply subtracting the chi-square and the degrees of freedom for the less restrictive model from the corresponding values for the more restrictive model:

$\chi^2_{diff} = 77.94 - 76.01 = 1.93$ with 70-69=1 degree of freedom. This statistic is not significant, which can be interpreted to mean that omitting the path from V7 to F5 does not lead to a significant worsening of fit. This test is another way to evaluate parameter significance.

If the model contains a lot of parameters, and especially if the researcher's goals are largely exploratory, conducting chi-square difference tests for many parameters may be impractical, because many additional models may need to be run. To assist with this problem, EQS prints *Wald tests* for dropping parameters (Chou & Bentler, 2002), which provide estimates of the expected gain in the chi-square value if various paths were dropped from the model. Wald tests complement LM tests in this way. In the EQS output, both multivariate and univariate Wald tests are listed. We see that we can drop the two nonsignificant paths from V7 to F5 and from F4 to F5 without significantly affecting model fit. In fact, the multivariate test tells us that we can drop *both* of them. To illustrate the non-equivalence of the z-test and the Wald test (or the chi-square difference test), note that not all insignificant paths are listed. For example, the correlation between V7 and F4, which was not significant according to the z-statistic, is not listed, suggesting that if we omitted this correlation from the model, we would significantly affect fit. When we do omit this path, we get the chi-square value of 80.78 with 70 degrees of freedom, and the resulting chi-square difference test of 4.77 with 1 degree of freedom, which is indeed significant.

Advanced Concepts

In this section, we discuss some of the more advanced topics in SEM. Due to space limitations, we can only cover selected topics that emphasize models for continuous variables. More complex methodology, such as that involving ordered categorical variable models (e.g., Lee, Poon, & Bentler, 1995; Muthén, 2001) or discrete choice models (e.g., Ashok, Dillon, & Yuan, 2002) is not discussed. We introduce statistical methods appropriate for nonnormal data and for incomplete data, as both these types of data are very common. We also describe some of the more advanced SEM models, such as growth curve models, multiple group models, and multilevel models. Finally we discuss an alternative to SEM called Partial Least Squares (PLS). Our discussion of these topics is necessarily brief.

Statistical Alternatives for Nonnormal Data

When data are not normally distributed, statistics based on normal theory can be misleading, but simple corrections are available to deal with this. Two standard measures of departure from normality are skewness and kurtosis. Skewness refers to the degree of asymmetry in the distribution. The normal distribution is symmetric and hence has zero skewness. Kurtosis is the relative ratio of the mass of the distribution located in the center versus in the tails. The kurtosis of the normal distribution is 3 and can be normed to be zero, so that "positive kurtosis" means that a distribution has heavier tails relative to the normal. Kurtosis is also a simple function of the fourth-order moments of the distribution (that is, "variances of the variances"). As we already mentioned, the ML parameter estimates are fairly robust to nonnormality, and have been shown to be less biased than estimates obtained via other methods that make fewer assumptions, such as AGLS, unless the sample size is huge. The ML test statistic and the standard errors, on the other hand, are affected by nonnormality, and in particular by positive kurtosis, which tends to inflate the test statistic (so that it is no longer chi-square distributed) and to lead to negative bias in the standard errors.

One solution is to adjust the test statistic by a scaling constant so that it approximates the mean

of the chi-square distribution more closely. That is, $T_{scaled} = \frac{1}{c} T_{ML}$. The appropriate scaling constant c is

a complicated function of the fourth order moments of the distribution, and will be greater than 1 when

the distribution is heavy-tailed. This scaled test statistic is also known as the Satorra-Bentler *scaled chi-*

square (Satorra & Bentler, 1994). Using the same information contained in fourth-order moments, we

can also compute *robust standard errors* that properly reflect additional variability in the estimates due

to nonnormality. This approach of using ML parameter estimates in conjunction with the scaled chi-

square and robust standard errors turns out to work rather well and is recommended over distribution-

free methods such as AGLS.

In EQS, these adjustments can be obtained by specifying METHOD=ML, ROBUST. Of course,

before we do that, we might want to find out whether we should even be concerned about nonnormality.

EQS includes univariate estimates of skewness and kurtosis as part of the descriptive statistics portion of

any output (this was omitted from Table 17.5). A more useful statistic is a measure of *multivariate*

kurtosis, and especially its normalized estimate, which works like a z-statistic. From Table 17.4, we see

that the normalized estimate in our dataset is about 6.9, which is greater than 1.96, a standard cut-off

value for a z-score. In practice only z-scores larger than about 3 are considered worrisome. Because our

modified model passed the ML chi-square test, the Satorra-Bentler chi-square, which is nearly always

lower, will not provide new information. Robust standard errors, on the hand, might differ enough from

the ML standard errors as to affect the significance of some of the parameters. In the EQS output, the

robust standard errors will be listed below the usual standard errors.

However, we choose a more interesting illustration and run METHOD=ML, ROBUST on the

original, unmodified model. It is possible that the robust test statistic will be sufficiently low as to not

require any modifications. The goodness of fit summary for this run is given in Table 17.5. The Satorra-

Bentler chi-square is 97.38 with 70 degrees of freedom and the associated p-value of 0.02, and we would still reject our original model. We can also see from the output that robust versions of some of the fit indices are available, and, in general, are not equal to their ML counterparts. Finally, a few other statistics are printed below the scaled chi-square; these are alternative and relatively new tests of model fit for nonnormal data. Of these, the F-statistic holds the most promise and has been found to perform as well as or better than the Satorra-Bentler chi-square in limited simulations. The p-value for it in our run is 0.06, which suggests that the model could be retained, but only marginally.

[TABLE 17.5 ABOUT HERE]

Handling Missing Data

Estimation methods we have considered thus far assume that the data are complete. However, many real life datasets contain missing data. If the proportion of missing data is small (say, below 5%), we can compute the sample covariance matrix S using only the complete cases without much loss of information. This approach is known as *listwise* deletion and is activated by typing MISSING=LISTWISE; in the Specifications section of the EQS input file. An alternative way is to employ *pairwise* deletion, which would compute each element of the sample covariance matrix S using all the data available for the computation of that element. This approach is activated by typing MISSING=PAIRWISE. This method can result in a sample covariance matrix that lacks consistency (the formal term for this is that it may fail to be positive definite); an extreme example is that a correlation estimate may exceed 1.0.

Before we present a better approach, we make some assumptions about how the missing data came about. We assume that whether a value is observed or missing is only a function of other observed quantities in the data, and not of the missing value itself (the accepted term for this is MAR, or *missing at random*). This definition of course allows that missingness be completely independent of all the

variables in the dataset (a special case known as MCAR, or *missing completely at random*). The approach to handling missing data we recommend is based on the assumption of multivariate normality. While this seems restrictive, corrections similar to the Satorra-Bentler corrections for complete data exist and can be applied once the parameter estimates have been obtained. This method of obtaining parameter estimates is known under multiples names, some of which are direct ML, FIML (full information ML), or raw data ML. This method is simply the extension of the ML estimation method to incomplete data and it obtains parameter estimates that maximize the probability of observed *incomplete* data under the model, assuming an ignorable missingness mechanism (MCAR or MAR). In EQS, this approach is activated by typing MISSING=ML;. If the data are nonnormal as well as incomplete, specifying robust estimation in addition to this command will produce the scaled chi-square for incomplete data, known as the Yuan-Bentler chi-square, and robust standard errors (Yuan & Bentler, 2000).

Growth Curve Models

The type of models we have discussed thus far are also known as *covariance structure models*, because they test the basic hypothesis that $\Sigma = \Sigma(\theta)$. It is also possible to test hypotheses about *mean structures*, which are of the form $\mu = \mu(\theta)$. *Growth curve models* are mean and covariance structure models that are appropriate for longitudinal data, in particular if trends over time are of interest. Suppose we collect data on four different occasions and are interested in whether they display an increasing trend. For example, the data can be scores on a math quiz collected over four weeks, and we hope to see an improvement in scores from one testing occasion to the next. For any given student, we can plot the four scores to obtain an individual trajectory, or a *growth curve*, representing his or her progress over time. If we have a group of students, we can represent their quiz scores at each time point as separate variables (say, V_1, V_2, V_3, and V_4). The means of these variables (μ_1, μ_2, μ_3, and μ_4) represent the

average weekly performance for the group, and the variances represent how much individual students'

trajectories vary from the overall group trajectory at each time point.

The hypothesis about a linear improvement in math scores can then be translated into a

hypothesis about a mean structure: $(\mu_1, \mu_2, \mu_3, \mu_4) = (a, a+b, a+2b, a+3b)$, where a and b are the

intercept and the slope. In other words, a represents the average quiz score at the initial time point, and

b represents the average weekly improvement for the group. But there is variability, from student to

student, in both the starting point and in the weekly increase. To capture this variability, we can create a

latent factor labeled Intercept, whose mean is a, and a latent factor labeled Slope, whose mean is b.

The variances of these latent factors reflect the variability of individual students' starting points and

growth rates. Figure 17.6 illustrates this idea. An unusual feature of this model is that the factor loadings

are fixed, as they represent the coefficients in front a and b in the mean structure hypothesis (1, 1, 1, 1;

and 0, 1, 2, 3), but the means (a and b) and the variances of the latent factors are freely estimated. The

means of the latent factors are represented in the diagram via paths from a new variable labeled Constant

to the latent factors; such a representation may be a bit confusing but is just a trick to get the means into

the diagram. It is also standard to allow the Intercept and the Slope to correlate, and this correlation is

usually negative. In our example, those students with higher initial quiz scores might have less room to

improve. Our final observation is that while Figure 17.6 arose from a hypothesis about a mean structure,

the equations implied by this diagram also impose a covariance structure on the data, although

explicating it is not especially enlightening.

[FIGURE 17.6 ABOUT HERE]

More complicated growth curve models often include covariates or variables influenced by the

growth factors. For example, using the ideas of our earlier figures, it would be easy to add a variable

(such as studying habits) that would predict the initial starting point and the rate of growth. See Duncan et al. (1999) for a more detailed introduction to growth curve models.

Multiple Group Models

Multiple group models are useful when we want to know whether the same model holds in different populations, or whether there is an interaction between group membership and any aspect of a structural model. For example, we can ask whether the TPB model of dieting behavior holds for men and women alike. In practice, this would mean having two data sets, which we assume are independent, and two model specifications as in Table 17.2. Even if the entire model structure is not identical across groups, it may be that certain important parameters are invariant across groups. It may be that intentions to diet are better predicted for one group than for another, or that they are predicted by different factors. Comparison of groups is not limited to covariance structure models. Returning to our linear growth curve model, we can compare the trajectory of children's improvement in mathematics separately for boys and girls. We can ask, for example, whether both groups start out at the same level of achievement at week 1, which translates into asking whether the mean of the Intercept factor is the same for the two groups.

Running a multiple group model is not the same as running separate SEM analyses for each group and comparing the results. When a multiple group model is run, the estimation for all the groups proceeds simultaneously and only one chi-square statistic is printed. This statistic can be an instance of the many types of statistics that are available for a single group, such as ML, AGLS, or ML, ROBUST. We must run the analysis for both groups simultaneously so that we can constrain parameters of interest to be equal across groups, allowing us to test hypotheses of *invariance*. While any free parameter, or any set of free parameters, can be evaluated for invariance or equality across populations, certain types of parameters tend to be evaluated

together, ideally in a sequence of nested model comparisons. Returning to the example of the

TPB model estimated as a multiple group model for men and women, we can first specify that all

the factor loadings for Cognition, Affect, PBC, and Norms be equal across the two groups. If the

resulting chi-square statistic is not significant, we can conclude that the model fits the data for

both groups, and furthermore, that we have *measurement invariance*. Even this step is not always

easily achieved, and the concept of *partial measurement invariance* also exists (e.g., Byrne,

Shavelson, & Muthén, 1989). A more specific test of invariance involves the chi-square

difference test between the model with all factor loadings free, and the model with them

constrained equal. We do not have to stop here and can ask whether structural paths are also

invariant across groups by conducting a chi-square difference test between the current model and

the model with further constraints imposed on the structural parameters. Finally, we can

constrain factor variances and covariances to be equal across groups, and as the last step, the

error variances. If this operation is successful, we will have achieved full invariance: that is, the

model is identical in both groups and can be tested on the sample as a whole.

Multilevel Models

Marketing and consumer data often have a hierarchical, or multilevel, structure. For example,

organizations are sampled first, and then individuals are chosen from within those organizations. There are

two main problems with ignoring the hierarchical aspects of such data, namely a single model

covariance matrix Σ may mix two unrelated sources of variance, and the statistical analyses may violate

assumptions since respondents are no longer independent. As in analysis of variance, the key idea of

multilevel models is to partition the total variance into component parts. The component parts are the level-

1 data (respondents) and the level-2 data (organizations). If there are a sufficient number of level-2 units, it

is possible to separate the total model covariance matrix into a between part (between organizations) and

within part (within respondents), that is, $\Sigma = \Sigma_B + \Sigma_W$. Muthén and Satorra (1995) refer to the overall

analysis of Σ as aggregated modeling, and the analysis of both Σ_W and Σ_B as disaggregated modeling,

which provides the most information. Σ_W gives the covariation at the respondent level, controlling for

variation in organizations, and Σ_B gives covariation at the level of the organization. If the organizational

differences do not matter, then Σ_B is zero, and $\Sigma = \Sigma_W$. Unfortunately, this is rare.

Since there are now two model matrices, multilevel models are handled like multiple group

models, although the multigroup assumption of independent groups is not used. The within model is set

up in one group, and the between model is set up in a second group. The types of models that can be

entertained at each level are any of the models discussed previously, though if there are mean structures,

these are used only for the between-model to explain mean differences across organizations. The

optimal estimation method is ML (Liang & Bentler, 2004) when the data are normal at each level.

Corrections for nonnormality also exist (Yuan & Bentler, 2003) and are available in EQS with the

METHOD=ML,ROBUST command.

When developed in a regression framework, multilevel modeling is often known as hierarchical

linear modeling (see Chapter 14, Advanced Regression Models), mixed model regression, random

effects, or random coefficient modeling. Such models contain a regression equation at each level, and

the coefficients at one level may be dependent variables in equations at another level. These models can

be estimated with specialized software (e.g., Raudenbush & Bryk, 2002).

Partial Least Squares Modeling

As structural equation modeling was being developed, Wold (1975, 1982) expressed reservations

about the strong assumptions, such as the existence of a precisely correct model and multivariate

normality of variables. He developed a method of "soft modeling" that would be more widely applicable

in fields where information is complex, variables are numerous, and cases or observations are minimal.

This approach, known as partial least squares (PLS), involves a series of approximate least squares operations. The model underlying PLS is the same as the SEM model, but estimation methods differ. However, standard SEM programs cannot do PLS estimation. An exposition of the contrast between PLS and ML was given in Jöreskog and Wold (1982).

A key feature of the PLS approach is that factors are approximated by weighted sums of observed variables. That is, "factor scores" are obtained, although not with the usual factor score formulae (see Chapter 18). Taking each factor in turn, a given factor is approximated by a weighted combination of only those variables that are indicated by it, giving a so-called "outside" approximation of each factor. Then each such factor score is predicted by other factor scores that are directly connected to the factor using -1 or +1 weights, depending on the sign of interfactor correlation, giving a new "inside" approximation of the factor scores. Then new estimates for the measurement model weights are obtained by multiple regression, regressing observed on inside factor scores (mode B), or simple regression of inside factor scores on observed variables (mode A). The entire process is repeated until convergence of the weights. Once this is done and optimal factor scores and weights are obtained and fixed, the actual path and loading coefficients of the model are estimated by ordinary least squares regression. Finally, means are estimated, if needed.

A serious problem with the PLS approach is that factor scores and factors are not identical, and regressions on factor scores generally are biased as estimators of regressions among latent variables. The approach does possess "consistency at large," meaning that as the number of indicators of a given latent variable approaches infinity, this bias will disappear; but most studies have a limited number of indicators per factor. In addition, there is no statistical test of the overall model fit, nor are there standard errors of parameter estimates, so that it is hard to evaluate model fit. The latter problems can be solved to some extent by bootstrapping.

The critique that Wold had of SEM methodology was largely directed at his suspicion about normal theory ML. Because methods for dealing with nonnormality in SEM now exist, we would recommend these newer SEM methods over the use of PLS, unless the model is based on a huge number of variables (even after parceling) and relatively few cases. In such situations, PLS may be the only practical method.

The Do's and Don'ts of SEM

SEM is a powerful statistical technique when used appropriately. But the theory behind SEM is fairly technical, and the modeling process can be frustrating and complicated. The result is that this technique is often misused. Below we list some potential pitfalls.

Ignoring test of model fit

Often, quite the opposite is considered a pitfall. That is, it is said that tests of model fit are sensitive to sample size and will reject models even with trivial misspecifications (e.g., Kline, 2005). While this is true, we find that this argument is way overused and often applied to sample sizes of just a few hundred. "Trivial misspecifications" refers to the size of the standardized residuals, which can be directly examined and reported to illustrate that they are in fact trivial, but this is almost never done. *Don't:* Blame the sensitivity of the test statistic unless your sample size is more than a few hundred and you have examined and reported your largest standardized residuals.

Do: Consider the most appropriate estimation method and test of model fit, which will deal with the problems of missing data and nonnormality to provide you with a reliable test statistic.

Basing Model Acceptance or Rejection on One or Two Global Fit Indices

Related to the previous point, often the researcher ignores the test of fit and the residuals and considers only indices such as the CFI. In fact, residuals are probably the single most informative set of fit indices because they point directly to the location and the size of the discrepancy (in meaningful

units) between the observed and the expected covariances. In addition, cut-off values for global fit indices are largely arbitrary and are designed as rough guidelines only. A model with a low CFI is probably in need of modification, but high CFI can be argued to be less informative as it can mask problems as well.

Don't: Have a "fit index tunnel vision" (Kline, 2005, p. 321).

Do: Consider multiple fit indices and examine and report the residuals.

Frowning on Model Modification

Often model modification procedures, such as adding and removing paths, are said to capitalize on chance, meaning that they extract relationships unique to the specific sample at hand. Of course this is true, and if model modification is done, cross validation on a new sample is recommended. But correlations based on even a few hundred observations are probably pretty stable, and model modification can then be a useful tool for discovering the set of relationships that can explain the observed covariances.

Don't: Dismiss model modification unless you truly have a small sample.

Do: Discover what your data have to say. Of course, be honest and report that related results are post-hoc and require cross-validation if possible.

Going Wild with Model Modification

On the flipside, it is important to approach model modification sensibly. Modifications that cannot be explained theoretically or are done solely for the purpose of improving fit tell us nothing. Correlated errors are hard to explain, but if you add them, you should make an effort to do so (what do these two items share in addition to being indicators of the same factor?). Cross-loadings, i.e., allowing one item to be indicator of more than one factor, are even more of a challenge. In many areas they are logical, but if you add these and can't explain them, you're essentially admitting you don't know what is

going on in your data. And indeed, if the sample size is small, results may not hold up (e.g., MacCallum, Roznowski, & Necowitz, 1992).

Don't: Let the LM tests drive your modifications without consulting with you first.

Do: Make sure the modified model is consistent with some theory, even if it is not the original theory.

Inferring Causation and Global Truth

The logic of SEM goes like this: If we assume that the specified set of causal relationships generated the data at hand, then the covariance matrix of the data must have a certain structure. If we refute the hypothesis that it has the given structure, we refute the assumption that the specified set of causal relationships generated the data. Unfortunately, failing to reject the hypothesis of the specified structure (the desired outcome unless you are examining a rival theory) does not imply that the specified set of causal relationships holds. In particular, in many models, the causal arrows can be reversed without affecting fit (see e.g., Hershberger, 1994; MacCallum et al., 1993). More globally, a completely different structure may exist that also explains the observed covariances.

Don't: Draw causal conclusions from correlational data simply because you used SEM.

Do: Remember causality can only be addressed via design, and that other well-fitting models to the same data may exist.

Equating R^2 and a Well-fitting Model

Overall model fit and R^2 values for various equations provide very different information about a model. The overall model may fit well, or badly, and there may be some large, or small, R^2. As a result, both sources of information need to be looked at to evaluate a model. If a theory implies a huge R^2 for some dependent variable, but it is only trivial in size, then even if the overall model fits the data, that aspect of the theory has been invalidated.

Don't: Assume your constructs are strongly related because your model fits.

Do: Remember to be concerned about the actual size of the relationships within the model before asserting that the theory holds.

Estimating a Complex Model on a Small Sample

SEM is a large sample technique. A popular rule of thumb is to have at least 10 observations per variable. If a large model is tested with a sample size of 100, estimation problems are likely, and additionally test statistics have extremely low power. We are pushing it in our dieting example. We employ item parceling to partially alleviate this problem. We also make sure we have enough power to at least reject the independence model.

Don't: Trust conclusions drawn from a model that is based on an extremely small sample.

Do: Simplify the model if necessary. Plan ahead of time and collect more data.

Books

The following recent books provide readable overviews of structural modeling: Kline (2005), Loehlin (2004), and Schumacker and Lomax (2004).

SEM Programs

The following web pages provide links to the most widely used SEM programs. All but Mx are commercial programs.

- AMOS http://www.spss.com/amos/

- PROC CALIS (available in SAS http://support.sas.com/software/index.htm)

- EQS http://www.mvsoft.com/

- LISREL http://www.ssicentral.com/lisrel/mainlis.htm

- Mplus http://www.statmodel.com/index2.html

- Mx http://www.vcu.edu/mx/

- RAMONA (available in SYSTAT http://www.systat.com/products/Systat)

- SEPATH (available in Statistica www.statsoft.com)

ENDNOTES

[1]This work was supported in part by National Institute of Drug Abuse grants DA01070 and DA00017.

[2]We thank Professor Richard Bagozzi for kindly providing this data set.

[3]The rules of covariance algebra are very simple: cov(X,Y+Z)=cov(X,Y)+cov(X,Z) and cov(bX)=bcov(X), where b is a constant.

[4]We discuss certain constraints in the context of multiple groups below.

[5]The procedure of first doing a factor analysis, then estimating the factor scores, and finally computing regressions among factor scores, or doing a path analysis with those factor scores, is also sometimes done to reduce model complexity. Unfortunately, such regressions do not provide unbiased estimates of regressions among the true factors.

[6]EQS 6.1 also allows the user to input the model in diagram form, which it then converts to appropriate syntax.

[7]A general equation for the WLS fitting function in matrix form is $F = (s - \hat{\sigma})'W(s - \hat{\sigma})$, where s is the vector of unique elements of S, $\hat{\sigma}$ is the vector of the unique elements of $\hat{\Sigma}$, and W is the matrix of weights.

[8]$F_{ML} = \ln|\Sigma| + tr(S\Sigma^{-1}) - \ln|S| - p$

[9]Hu and Bentler (1999) recommend a less stringent SRMR<0.08.

[10]Multivariate LM tests are available, but because they automatically choose the largest path at each step, nonsensical (from the point of view of theory) models often result. Even the Univariate LM tests can give questionable results – the predicted standardized coefficient of 4.497 for F1→F5 is unreasonably large, implying that this would not be a stable path to add.

[11]This question can also be asked when the original model does not fit the data, but then the parameter tests may not be meaningful (Yuan & Bentler, 2004).

REFERENCES

Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes*, *50*, 179-211.

Ashok, K., Dillon, W. R., & Yuan, S. (2002). Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research*, *39*, 31-46.

Bagozzi, R. P. (1980). *Causal models in marketing*. New York: Wiley.

-----, & Warshaw, P. R. (1990). Trying to consume. *Journal of Consumer Research, 17*(2), 127-140.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumaker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum Associates.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing, 13*, 139-161.

Bentler, P.M. (2005). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

-----, & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, *45*, 289-308.

Bodur, H. O., Brinberg, D., & Coupey, E. (2000). Belief, affect, and attitude: Alternative models of the determinants of attitude. *Journal of Consumer Psychology, 9*(1), 17-28.

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

-----. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605-634.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows.* Thousand Oaks, CA: Sage.

-----, Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466.

Chou, C.-P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange Multiplier, and Wald tests. *Multivariate Behavioral Research*, *25*, 115-136.

-----, & -----. (2002). Model modification in structural equation modeling by imposing constraints. *Computational Statistics & Data Analysis*, *41*, 271-287.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Erlbaum.

Eagly, A. H., Mladinic, A., & Otto, S. (1994). Cognitive and affective bases of attitudes towards social groups and social policies. *Journal of Experimental Social Psychology, 30*, 113-137.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

-----, & Middlestadt, S. E. (1995). Noncognitive effects on attitude formation and change: Fact or artifact? *Journal of Consumer Psychology*, *4*, 181-202.

Hershberger, S. L. (1994). The specification of equivalent models before the collection of data. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 68-108). Thousand Oaks, CA: Sage.

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

Jöreskog, K. G., & Sörbom, D. (1994). *LISREL 8 user's reference guide*. Mooresville, IN: Scientific Software.

-----, & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables: Historical and comparative aspects. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 263-270). Amsterdam: North- Holland.

Kim, J., Lim, J.-S., & Bhargava, M. (1998). The role of affect in attitude formation: A classical conditioning approach. *Journal of the Academy of Marketing Science, 26*(2), 143-152.

Kline, R. B. (2005). *Principles and practice of structural equation modeling.* New York: Guilford.

Lee, S.–Y., Poon, W.–Y., & Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology*, *48*, 339-358.

Liang, J., & Bentler, P. M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika*, 69, 101-122.

Little, T. D., Cunningham, W., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighting the merits. *Structural Equation Modeling, 9*, 151-173.

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis*. Mahwah, NJ: Erlbaum.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201-226.

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

-----, Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185-199.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variables. *Psychological Methods*, *7*, 83-104.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers of overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 291-322). Washington, DC: American Psychological Association.

-----, & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. V. Marsden (Ed.), *Sociological methodology 1995* (pp. 267-316). Cambridge, MA: Blackwell.

Perugini, M., & Bagozzi, R. (2001). The role of desires and anticipated emotions in goal-directed behaviours: Broadening and deepening the theory of planned behaviour. *British Journal of Social Psychology, 40*, 79-98.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed.). Thousand Oaks: Sage.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd Ed.). Mahwah, NJ: Erlbaum.

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*, 371-384.

Steenkamp, J.-B. E. M., & Baumgartner, H. (2000). On the use of structural equation models for marketing modeling. *International Journal of Research in Marketing, 17*, 195-202.

Trafimow, D., & Sheeran, P. (1998). Some tests of the distinction between cognitive and affective beliefs. *Journal of Experimental Social Psychology, 34*, 378-397.

Wold, H. (1975). Path models with latent variables: The NIPALS approach. In H. M. Blalock Jr., A.

    Aganbegian, F. M. Borodkin, R. Boudon, & V. Capecchi (Eds.), *Quantitative sociology* (pp. 307-357).

    New York: Seminar.

-----. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog & H. Wold (Eds.)

    Systems under indirect observation: Causality, structure, prediction (Part II, pp. 1-54). Amsterdam:

    North- Holland.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance

    structure analysis with nonnormal missing data. In *Sociological methodology 2000* (pp. 165-200).

    Washington, DC: American Sociological Association.

-----, & -----. (2003). Eight test statistics for multilevel structural equation models. *Computational*

    Statistics & Data Analysis, *44*, 89-107.

-----, & -----. (2004). On chi-square difference and z tests in mean and covariance structure analysis

    when the base model is misspecified. *Educational and Psychological Measurement*, *64*, 737-757.

Table 17.1

The Correlation/Covariance Matrix of the 16 Variables in the Dieting Data Set

	COGN1	COGN2	COGN3	AFF1	AFF2	AFF3	NORM1	NORM2	NORM3	PBC1	PBC2	PBC3	INT1	INT2	INT3	BEH
COGN1	1.73	0.66	0.63	0.42	0.29	0.29	0.24	0.29	0.19	0.37	0.04	0.07	0.64	0.64	0.60	0.41
COGN2	1.11	1.62	0.81	0.60	0.44	0.51	0.33	0.37	0.19	0.40	0.09	-0.03	0.43	0.46	0.44	0.38
COGN3	1.05	1.30	1.60	0.62	0.50	0.53	0.29	0.28	0.17	0.41	0.07	0.02	0.44	0.45	0.44	0.32
AFF1	0.68	0.93	0.96	1.51	0.74	0.82	0.34	0.32	0.13	0.38	0.27	0.12	0.27	0.26	0.25	0.25
AFF2	0.51	0.75	0.84	1.20	1.78	0.76	0.33	0.18	0.12	0.42	0.29	0.18	0.26	0.226	0.22	0.20
AFF3	0.58	1.00	1.04	1.55	1.56	2.38	0.33	0.28	0.11	0.36	0.27	0.12	0.23	0.25	0.26	0.21
NORM1	0.74	0.99	0.85	0.96	1.02	1.17	5.45	0.55	0.39	0.13	0.13	0.11	0.22	0.30	0.24	0.33
NORM2	0.81	0.99	1.04	0.82	0.51	0.91	2.70	4.45	0.40	0.24	0.24	0.14	0.15	0.27	0.17	0.16
NORM3	0.51	0.50	0.85	0.33	0.33	0.34	1.88	1.76	4.28	0.06	0.11	0.04	0.22	0.29	0.29	0.16
PBC1	1.15	1.21	0.75	1.10	1.33	1.30	0.73	1.20	0.29	5.58	0.59	0.34	0.36	0.40	0.24	0.21
PBC2	0.31	0.30	0.45	0.83	0.97	1.04	0.78	1.25	0.57	3.53	6.35	0.75	0.07	0.17	0.01	-0.05
PBC3	0.25	-0.10	1.22	0.39	0.64	0.47	0.69	0.80	0.23	2.16	5.00	7.04	-0.01	0.02	-0.12	-0.03
INT1	1.54	1.00	0.22	0.60	0.62	0.65	0.95	0.59	0.83	1.57	0.33	-0.06	3.32	0.75	0.63	0.35
INT2	1.39	0.97	0.08	0.54	0.58	0.64	1.16	0.94	1.00	1.56	0.70	0.07	2.26	2.74	0.77	0.38
INT3	1.50	1.06	1.06	0.58	0.55	0.76	1.09	0.68	1.15	1.07	0.04	-0.60	2.19	2.43	3.62	0.29
BEH	0.52	0.46	0.39	0.29	0.26	0.31	0.74	0.33	0.32	0.48	-0.11	-0.07	0.61	0.61	0.54	0.93

Covariances are below the diagonal, correlations are above the diagonal, and variances are on the diagonal.

Table 17.2

Sample EQS Input for the Model in Figure 17.4

```
/TITLE
Model in Figure 17.4
/SPECIFICATIONS
DATA='data.ess';
VARIABLES=16; CASES=108; METHOD=ML; ANALYSIS=COVARIANCE; MATRIX=RAW;
/LABELS
V1=COGN1; V2=COGN2; V3=COGN3; V4=AFF1; V5=AFF2; V6=AFF3; V7=PBC1; V8=PBC2;
V9=PBC3; V10=NORS1; V11=NORS2; V12=NORS3; V13=BEH; V14=INT1; V15=INT2;
V16=INT3; F1=COGN; F2=AFF; F3=ATT; F4=NORMS; F5=INTENT;
/EQUATIONS
V1 = 1F1 + E1;
V2 = *F1 + E2;
V3 = *F1 + E3;
V4 = 1F2 + E4;
V5 = *F2 + E5;
V6 = *F2 + E6;
V10 = 1F4 + E10;
V11 = *F4 + E11;
V12 = *F4 + E12;
V13 = *F5 + E13;
V14 = 1F5 + E14;
V15 = *F5 + E15;
V16 = *F5 + E16;
F1 = 1F3 + D1;
F2 = *F3 + D2;
F5 = *F3 + *V7 + *F4 + D5;
/VARIANCES
 F3 = *;
 F4 = *;
 E1 = *;
 E2 = *;
 E3 = *;
 E4 = *;
 E5 = *;
 E6 = *;
 V7 = *;
 E10 = *;
 E11 = *;
 E12 = *;
 E13 = *;
 E14 = *;
 E15 = *;
 E16 = *;
 D1 = *;
 D2 = *;
 D5 = *;
/COVARIANCES
 F3,V7 = *;
 F3,F4 = *;
 V7,F4 = *;
/PRINT
FIT=ALL;
/LMTEST
PROCESS=SIMULTANEOUS;
SET=PVV,PFV,PFF,PEE,PDD,GVV,GVF,GFV,
GFF,BVF,BFV,BFF;
```

```
/WTEST
/END
```

Table 17.3

Selected EQS Output for the Syntax in Table 17.2

```
STANDARDIZED RESIDUAL MATRIX:


                   COGN1      COGN2      COGN3      AFF1       AFF2
                   V  1       V  2       V  3       V  4       V  5
       COGN1   V  1    .000
       COGN2   V  2   -.001       .000
       COGN3   V  3   -.019       .006       .000
        AFF1   V  4   -.029       .044       .073       .000
        AFF2   V  5   -.110      -.050       .011      -.014       .000
        AFF3   V  6   -.153      -.032       .002      -.003       .023
        PBC1   V  7    .020      -.029      -.014       .062       .141
       NORS1   V 10   -.025       .007      -.034       .095       .114
       NORS2   V 11    .026       .044      -.038       .079      -.033
       NORS3   V 12   -.006      -.046      -.060      -.043      -.036
         BEH   V 13    .229       .154       .099       .081       .054
        INT1   V 14    .296       .006       .024      -.046      -.024
        INT2   V 15    .224      -.047      -.046      -.109      -.069
        INT3   V 16    .246       .002       .013      -.070      -.069


                   AFF3       PBC1       NORS1      NORS2      NORS3
                   V  6       V  7       V 10       V 11       V 12
        AFF3   V  6    .000
        PBC1   V  7    .045       .000
       NORS1   V 10    .089      -.039       .000
       NORS2   V 11    .045       .070       .000       .000
       NORS3   V 12   -.065      -.065      -.008       .008       .000
         BEH   V 13    .048       .042       .198       .034       .065
        INT1   V 14   -.078       .040      -.024      -.092       .040
        INT2   V 15   -.115       .013       .005      -.023       .080
        INT3   V 16   -.054      -.092      -.009      -.081       .110


                   BEH        INT1       INT2       INT3
                   V 13       V 14       V 15       V 16
         BEH   V 13    .000
        INT1   V 14    .015       .000
        INT2   V 15   -.009      -.001       .000
        INT3   V 16   -.044      -.013       .007       .000


            AVERAGE ABSOLUTE STANDARDIZED RESIDUALS       =       .0495
   AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS   =       .0571


LARGEST STANDARDIZED RESIDUALS:

    NO.   PARAMETER   ESTIMATE     NO.   PARAMETER   ESTIMATE
    ---   ---------   --------     ---   ---------   --------
     1    V14, V1         .296     11    V5,  V1        -.110
     2    V16, V1         .246     12    V16, V12        .110
     3    V13, V1         .229     13    V15, V4        -.109
     4    V15, V1         .224     14    V13, V3         .099
     5    V13, V10        .198     15    V10, V4         .095
     6    V13, V2         .154     16    V16, V7        -.092
     7    V6,  V1        -.153     17    V14, V11       -.092
     8    V7,  V5         .141     18    V10, V6         .089
     9    V15, V6        -.115     19    V13, V4         .081
    10    V10, V5         .114     20    V16, V11       -.081


GOODNESS OF FIT SUMMARY FOR METHOD = ML
```

```
INDEPENDENCE MODEL CHI-SQUARE       =     912.729 ON    91 DEGREES OF FREEDOM

CHI-SQUARE =      107.569 BASED ON      70 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS        .00262


FIT INDICES
-----------
BENTLER-BONETT     NORMED FIT INDEX =      .882
BENTLER-BONETT NON-NORMED FIT INDEX =      .941
COMPARATIVE FIT INDEX (CFI)         =      .954
BOLLEN   (IFI) FIT INDEX            =      .955
MCDONALD (MFI) FIT INDEX            =      .840
LISREL   GFI  FIT INDEX             =      .885
LISREL   AGFI FIT INDEX             =      .827
ROOT MEAN-SQUARE RESIDUAL (RMR)     =      .190
STANDARDIZED RMR                    =      .075
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) = .071
90% CONFIDENCE INTERVAL OF RMSEA  ( .042, .096)


WALD TEST (FOR DROPPING PARAMETERS)
MULTIVARIATE WALD TEST BY SIMULTANEOUS PROCESS



           CUMULATIVE MULTIVARIATE STATISTICS           UNIVARIATE INCREMENT
         -----------------------------------          --------------------

STEP   PARAMETER   CHI-SQUARE  D.F. PROBABILITY      CHI-SQUARE  PROBABILITY
----   ---------   ----------  ---- -----------      ----------  -----------

  1     D1,D1         .401       1      .527             .401       .527
  2     F5,F4        2.298       2      .317            1.896       .168
  3     F5,V7        4.463       3      .216            2.166       .141


LAGRANGE MULTIPLIER TEST (FOR ADDING PARAMETERS)
ORDERED UNIVARIATE TEST STATISTICS:
                                           HANCOCK          STANDAR-
                                 CHI-       70 DF  PARAMETER  DIZED
  NO    CODE    PARAMETER       SQUARE  PROB. PROB.  CHANGE   CHANGE
  --   ------   ---------       ------  ----- ----- -------- ---------  --------
   1   2  21      F5,V1        29.162   .000  1.000     .762      .400
   2   2  20      V1,F5        27.124   .000  1.000     .455      .239
   3   2  21      F5,V4         7.782   .005  1.000    -.444     -.249
   4   2  22      F1,F5         7.652   .006  1.000     .444      .317
   5   2  22      F2,F5         7.652   .006  1.000    -.375     -.230
   6   2  10      D5,D1         7.652   .006  1.000     .457     1.528
   7   2  22      F5,F1         7.652   .006  1.000    6.295     4.497
   8   2  22      F5,F2         7.652   .006  1.000    -.589     -.362
   9   2  10      D5,D2         7.652   .006  1.000    -.385     -.429
  10   2  21      F1,V4         7.427   .006  1.000     .538      .454
```

Table 17.4

Selected EQS Output for the Modified Model in Figure 17.5

```
                        MULTIVARIATE KURTOSIS
                        ---------------------


 MARDIA'S COEFFICIENT (G2,P) =     28.2680
 NORMALIZED ESTIMATE =             6.9397


 CASE NUMBERS WITH LARGEST CONTRIBUTION TO NORMALIZED MULTIVARIATE KURTOSIS:
 -----------------------------------------------------------------------------

 CASE NUMBER          2         68          69          94          107

 ESTIMATE       314.4517    399.6403    244.9824    239.0308    191.7511


 LARGEST STANDARDIZED RESIDUALS:

     NO.   PARAMETER   ESTIMATE     NO.   PARAMETER   ESTIMATE
     ---   ---------   --------     ---   ---------   --------
      1    V13, V10       .197      11    V15, V12       .091
      2    V13, V2        .167      12    V14, V6       -.089
      3    V6,  V1       -.140      13    V10, V5        .087
      4    V16, V12       .115      14    V14, V11      -.085
      5    V15, V6       -.114      15    V12, V6       -.082
      6    V13, V3        .113      16    V13, V4        .075
      7    V7,  V5        .111      17    V16, V4       -.075
      8    V13, V1        .110      18    V11, V7        .075
      9    V15, V4       -.106      19    V16, V5       -.075
     10    V5,  V1       -.098      20    V14, V1        .072


 GOODNESS OF FIT SUMMARY FOR METHOD = ML

 INDEPENDENCE MODEL CHI-SQUARE        =    912.729 ON    91 DEGREES OF FREEDOM

 INDEPENDENCE AIC =    730.72936   INDEPENDENCE CAIC =   395.65542
        MODEL AIC =    -61.29229         MODEL CAIC = -315.35935

 CHI-SQUARE =       76.708 BASED ON      69 DEGREES OF FREEDOM
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS        .24527


 FIT INDICES
 -----------
 BENTLER-BONETT     NORMED FIT INDEX =      .916
 BENTLER-BONETT NON-NORMED FIT INDEX =      .988
 COMPARATIVE FIT INDEX (CFI)         =      .991
 BOLLEN   (IFI) FIT INDEX            =      .991
 MCDONALD (MFI) FIT INDEX            =      .965
 LISREL    GFI  FIT INDEX            =      .912
 LISREL   AGFI  FIT INDEX            =      .865
 ROOT MEAN-SQUARE RESIDUAL (RMR)     =      .154
 STANDARDIZED RMR                    =      .057
 ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)    =      .032
 90% CONFIDENCE INTERVAL OF RMSEA  (       .000,       .067)

 MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
 STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

 COGN1   =V1  =   1.000 F1    + 1.000 E1

 COGN2   =V2  =   1.250*F1    + 1.000 E2
                   .144
                  8.700@
```

```
COGN3   =V3  =   1.214*F1    + 1.000 E3
                  .141
                 8.601@

AFF1  =V4  =   1.000 F2    + 1.000 E4

AFF2  =V5  =    .972*F2    + 1.000 E5
                  .085
                11.438@

AFF3  =V6  =   1.236*F2    + 1.000 E6
                  .092
                13.496@

NORS1   =V10 =   1.000 F4    + 1.000 E10

NORS2   =V11 =    .867*F4    + 1.000 E11
                  .164
                 5.277@

NORS3   =V12 =    .624*F4    + 1.000 E12
                  .139
                 4.492@

BEH   =V13 =    .278*F5    + 1.000 E13
                  .063
                 4.393@

INT1  =V14 =   1.000 F5    + 1.000 E14

INT2  =V15 =   1.038*F5    + 1.000 E15
                  .094
                11.034@

INT3  =V16 =   1.049*F5    + 1.000 E16
                  .110
                 9.537@
```

CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

```
  COGN  =F1  =   1.000 F3    + 1.000 D1

   AFF  =F2  =    .973*F3    + 1.000 D2
                   .183
                  5.321@

 INTENT =F5  =    .067*V7   +  .812*F3   +  .113*F4   +  .735*E1
                   .058          .251          .094          .131
                  1.150         3.238@        1.203         5.619@

             + 1.000 D5
```

 VARIANCES OF INDEPENDENT VARIABLES

 STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

```
                  V                            F
                  ---                          ---
V7  -  PBC1              5.579*I F3  -  ATT          .721*I
                         .763 I                      .208 I
                        7.314@I                     3.461@I
                            I                            I
                            I F4  -NORMS            3.118*I
                            I                        .841 I
                            I                       3.708@I
                            I                            I


                  E                            D
                  ---                          ---
E1  -COGN1               .851*I D1  -  COGN          .137*I
                         .129 I                      .090 I
                        6.594@I                     1.523 I
                            I                            I
E2  -COGN2               .286*I D2  -  AFF           .578*I
                         .078 I                      .134 I
                        3.652@I                     4.314@I
                            I                            I
E3  -COGN3               .333*I D5  -INTENT          .893*I
```

```
                              .079 I                       .208 I
                             4.219@I                      4.293@I
                                 I                           I
     E4  -  AFF1            .247*I                           I
                             .064 I                           I
                            3.878@I                           I
                                 I                           I
     E5  -  AFF2            .584*I                           I
                             .097 I                           I
                            6.006@I                           I
                                 I                           I
     E6  -  AFF3            .457*I                           I
                             .104 I                           I
                            4.394@I                           I
                                 I                           I
     E10 -NORS1           2.328*I                           I
                             .596 I                           I
                            3.906@I                           I
                                 I                           I
     E11 -NORS2           2.109*I                           I
                             .479 I                           I
                            4.405@I                           I
                                 I                           I
     E12 -NORS3           3.065*I                           I
                             .481 I                           I
                            6.370@I                           I
                                 I                           I
     E13 -  BEH            .757*I                           I
                             .106 I                           I
                            7.170@I                           I
                                 I                           I
     E14 -  INT1          1.144*I                           I
                             .193 I                           I
                            5.931@I                           I
                                 I                           I
     E15 -  INT2           .387*I                           I
                             .121 I                           I
                            3.210@I                           I
                                 I                           I
     E16 -  INT3          1.216*I                           I
                             .207 I                           I
                            5.875@I                           I
                                 I                           I


    COVARIANCES AMONG INDEPENDENT VARIABLES
    ---------------------------------------
    STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.


                    V                           F
                    ---                         ---
    F3  -  ATT           1.041*I F4  -NORMS           .798*I
    V7  -  PBC1           .262 I F3  -  ATT           .231 I
                        3.973@I                      3.450@I
                             I                           I
    F4  -NORMS            .950*I                          I
    V7  -  PBC1           .485 I                           I
                        1.956 I                           I
                             I                           I


    STANDARDIZED SOLUTION:                                    R-SQUARED

    COGN1   =V1  =   .708 F1    + .706 E1                        .502
    COGN2   =V2  =   .908*F1    + .419 E2                        .824
    COGN3   =V3  =   .890*F1    + .457 E3                        .791
     AFF1   =V4  =   .914 F2    + .405 E4                        .836
     AFF2   =V5  =   .819*F2    + .573 E5                        .671
     AFF3   =V6  =   .899*F2    + .438 E6                        .808
    NORS1   =V10 =   .757 F4    + .654 E10                       .572
    NORS2   =V11 =   .725*F4    + .688 E11                       .526
    NORS3   =V12 =   .533*F4    + .846 E12                       .284
      BEH   =V13 =   .425*F5    + .905 E13                       .181
```

```
INT1  =V14 =   .809 F5    + .588 E14                                            .654
INT2  =V15 =   .926*F5   + .377 E15                                             .858
INT3  =V16 =   .814*F5   + .581 E16                                             .662
COGN  =F1  =   .917 F3    + .399 D1                                             .841
AFF   =F2  =   .736*F3   + .677 D2                                              .541
INTENT =F5 =   .107*V7   + .468*F3    + .136*F4    + .461*E1 + .642 D5 .588
```

```
CORRELATIONS AMONG INDEPENDENT VARIABLES
----------------------------------------
                    V                         F
                   ---                       ---
F3  - ATT               .519*I F4  -NORMS            .532*I
V7  - PBC1                   I F3  - ATT                  I
                            I                            I
F4  -NORMS              .228*I                            I
V7  - PBC1                  I                            I
                            I                            I
```

```
WALD TEST (FOR DROPPING PARAMETERS)
MULTIVARIATE WALD TEST BY SIMULTANEOUS PROCESS

          CUMULATIVE MULTIVARIATE STATISTICS          UNIVARIATE INCREMENT
          ----------------------------------          --------------------

STEP  PARAMETER   CHI-SQUARE  D.F. PROBABILITY    CHI-SQUARE  PROBABILITY
----  ----------  ----------  ---- -----------    ----------  -----------

  1   F5,V7         1.322      1      .250          1.322        .250
  2   F5,F4         2.270      2      .321           .948        .330
```

```
LAGRANGE MULTIPLIER TEST (FOR ADDING PARAMETERS)
ORDERED UNIVARIATE TEST STATISTICS:
                                         HANCOCK              STANDAR-
                            CHI-          69 DF  PARAMETER    DIZED
NO    CODE    PARAMETER     SQUARE  PROB. PROB.  CHANGE       CHANGE
--    ------  ---------     ------  ----- ------ ---------    --------
 1    2  21     F1,V4       9.403   .002  1.000     .557        .489
 2    2  22     F2,F5       6.550   .010  1.000    -.273       -.165
 3    2  20     V4,F1       6.500   .011  1.000     .303        .266
 4    2  21     F2,V15      6.269   .012  1.000    -.212       -.115
 5    2  21     F5,V4       5.786   .016  1.000    -.378       -.209
 6    2   6     E16,E15     5.301   .021  1.000     .416        .606
 7    2   6     E13,E10     4.923   .027  1.000     .348        .262
 8    2  21     F5,V13      4.705   .030  1.000    -.345       -.244
 9    2   6     E6,E1       4.582   .032  1.000    -.165       -.264
10    2  20     V13,F1      4.366   .037  1.000     .251        .282
```

Table 17.5

Selected EQS Output for the Model in Figure 17.4, with METHOD=ML, ROBUST

```
GOODNESS OF FIT SUMMARY FOR METHOD = ROBUST

ROBUST INDEPENDENCE MODEL CHI-SQUARE =      565.164 ON     91 DEGREES OF FREEDOM

SATORRA-BENTLER SCALED CHI-SQUARE =      97.3805 ON      70 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS         .01694

RESIDUAL-BASED TEST STATISTIC                         =    310.473
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS         .00000

YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC            =     80.127
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS         .19119

YUAN-BENTLER RESIDUAL-BASED F-STATISTIC     =      1.575
DEGREES OF FREEDOM   =                           70,    38
PROBABILITY VALUE FOR THE F-STATISTIC IS          .06449

FIT INDICES
-----------
BENTLER-BONETT      NORMED FIT INDEX =       .828
BENTLER-BONETT NON-NORMED FIT INDEX =       .925
COMPARATIVE FIT INDEX (CFI)          =       .942
BOLLEN   (IFI) FIT INDEX             =       .945
MCDONALD (MFI) FIT INDEX             =       .881
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)     =      .060
90% CONFIDENCE INTERVAL OF RMSEA  (        .027,        .087)
```
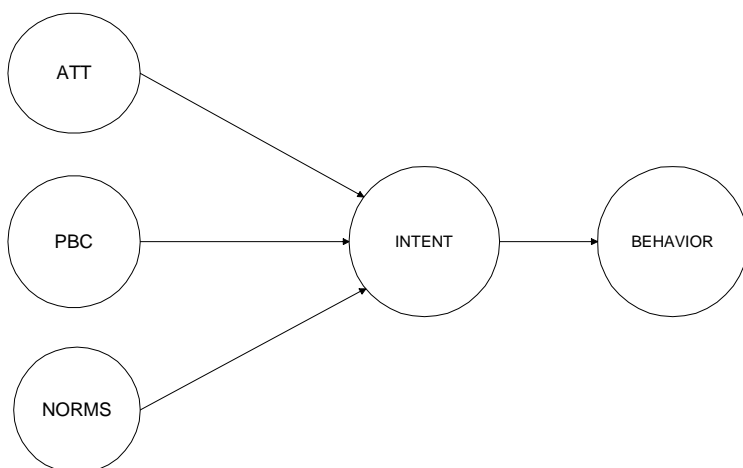
Figure 17.1

The Conceptual Model for TPB
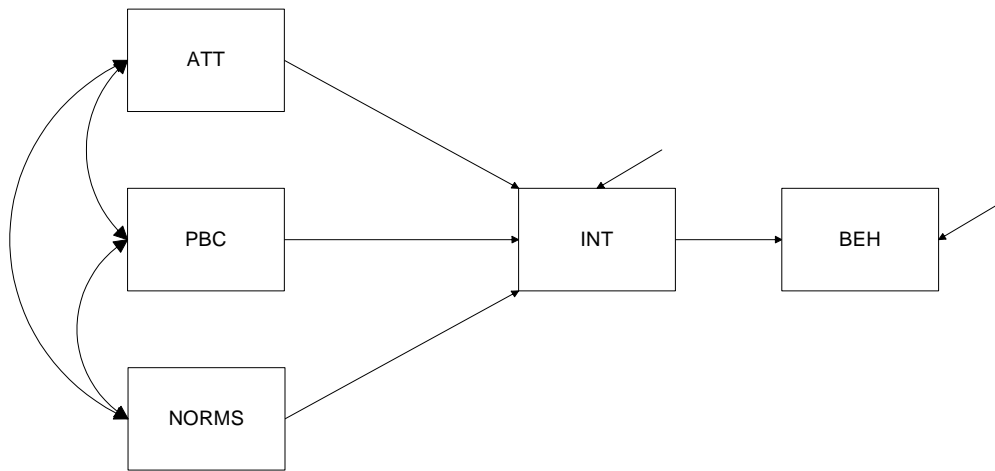
Figure 17.2

A Path Diagram for TPB

Figure 17.3

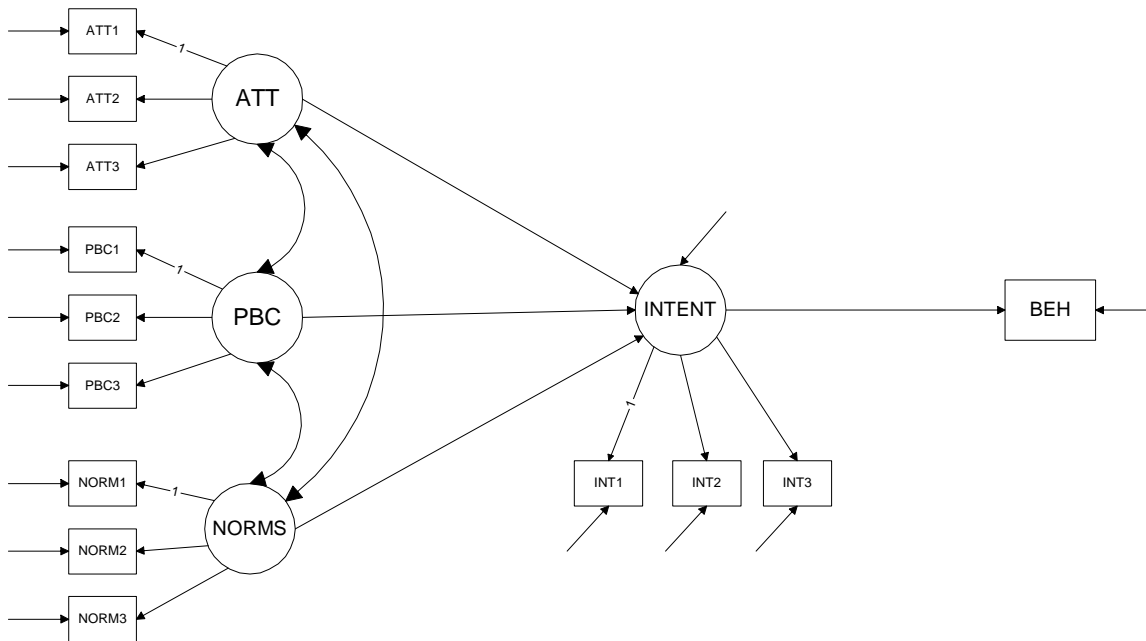**A Sample Structural Equation Model for TPB**

Figure 17.4

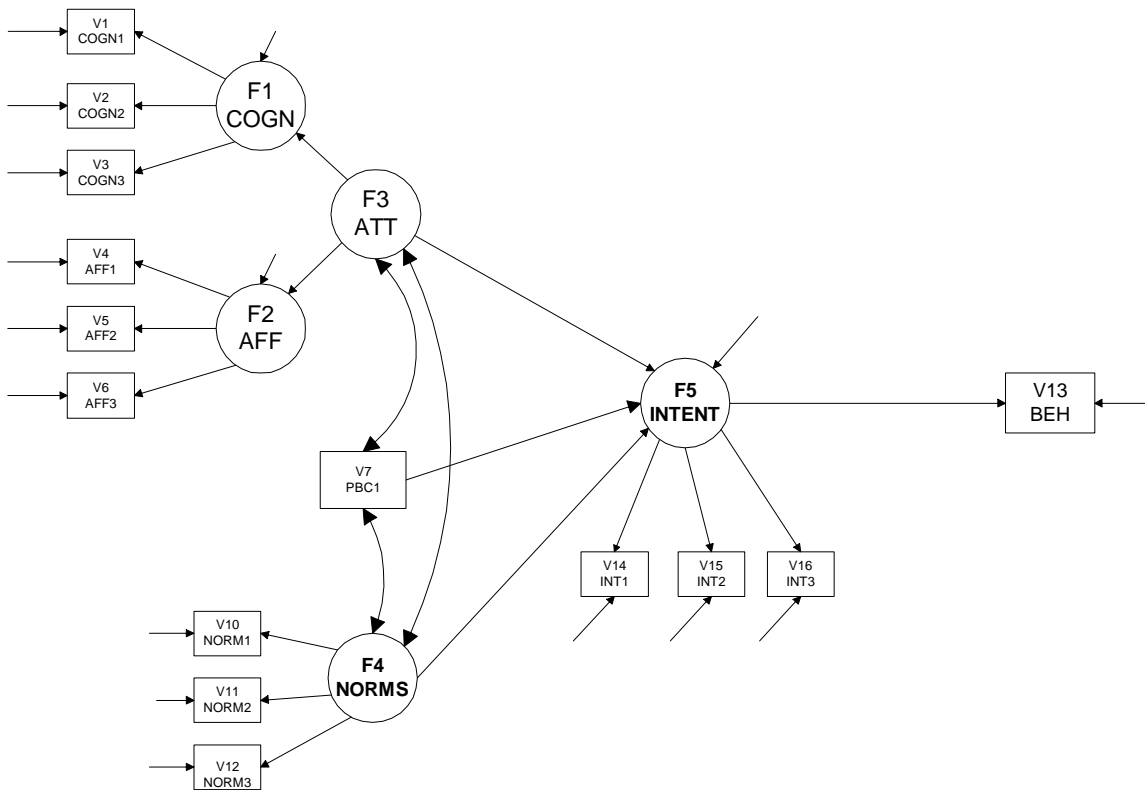**The Initial Tested Model**

Figure 17.5
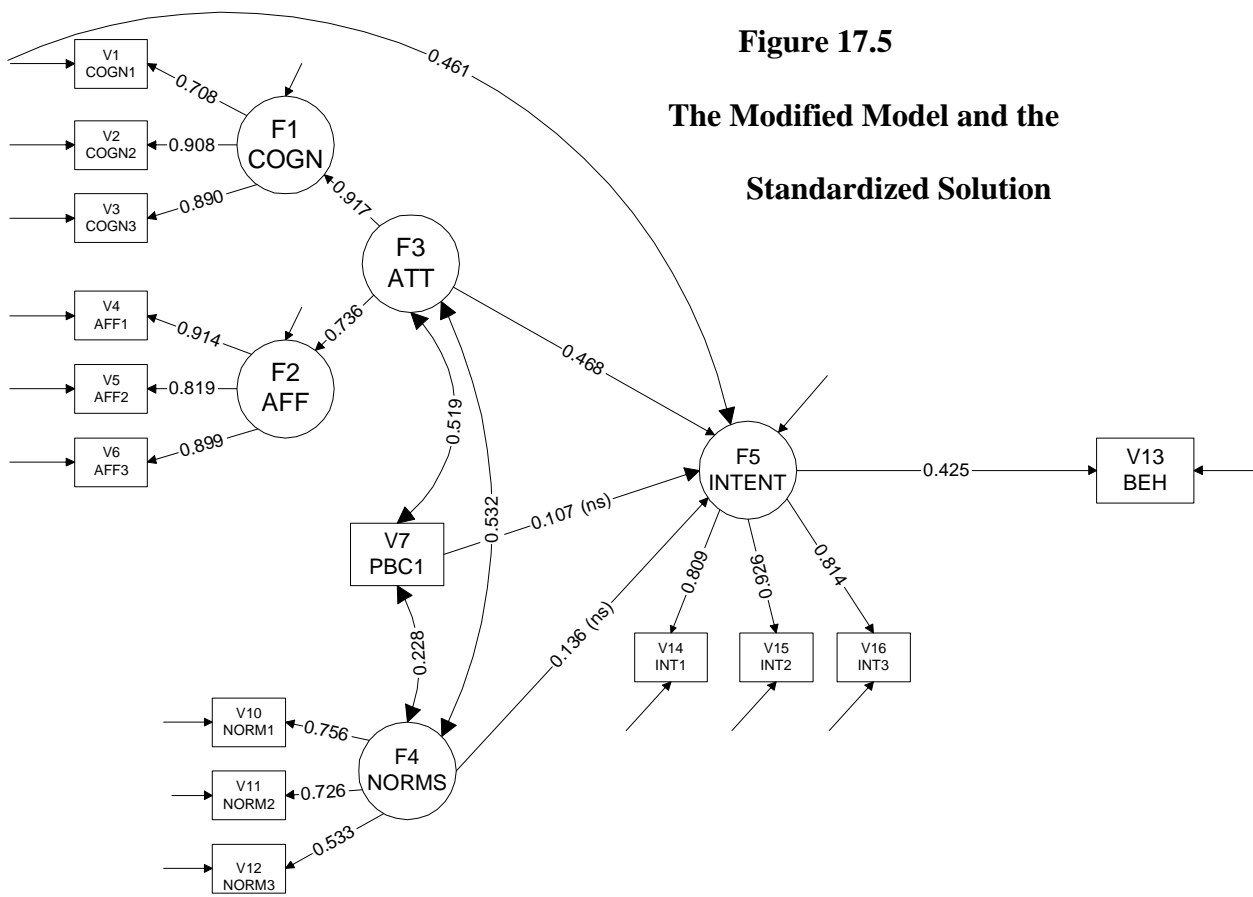
The Modified Model and the

Standardized Solution

Figure 17.6

A Linear Growth Curve Model with Four Time Points