1 Key Terms

Please review and learn these terms. SEM is extremely jargon-laden. I even recommend for this topic that you create flash-cards or vocabulary lists that you can memorize and refer to while studying. Below, new and jargon terms will be in italics. Please look them up if you need to. I will be defining these terms to some extent, but life will be easier for all of us if you have some exposure to them before lecture. My admonition to read the chapter before lecture is really crucial for this material.

Below are some of the more important terms you will need to be familiar with for this lecture:

**Causal inference** A dependence relationship of two or more variables in which the researcher clearly specifies that one or more variables is the *cause* of an outcome represented by at least one other variable.

**Causation** The principle by which cause and effect are established between two variables. It requires a sufficient degree of association (covariance), that the cause occurs before the effect, and that no other reasonable causes for the outcome are present. Strictly, causation is almost never proven, but in practice strong theoretical support can make empirical estimation of causation possible.

**Communality** The total amount of variance that a *measured* variable has in common with the *constructs* upon which it loads. Good measurement practice suggests that each measured variable should load on only one construct. Hence, it can be thought of as the variance explained in a measured variable by the construct. In *confirmatory factor analysis* (CFA), it is referred to as the squared multiple correlation for a measured variable.

**Confirmatory analysis** The use of a multivariate technique to test (confirm) a prespecified relationship. It is the opposite of *exploratory analysis*.

**Construct** An unobservable or *latent* concept that the researcher can define in conceptual terms but that cannot be directly measured (i.e. there is no single measurement that can be made that will totally and perfectly quantify the concept), or cannot be measured without error.
Endogenous constructs The latent, multivariate equivalents to dependent variables. It is represented by a variate of dependent variables.

Exogenous constructs The latent, multivariate equivalents to independent variables. They are constructs determined by factors outside the model.

Fixed parameter A parameter that has a value specified by the researcher. Most often the value specified is zero, indicating no relationship, although in some instances a non-zero value can be specified.

Free parameter A parameter estimated by the structural equation program to represent the strength of a specified relationship. These parameters may occur in the measurement model (most often denoting loadings of indicators to constructs), as well as in the structural model (relationships among constructs).

Indicator An observed value (also called a measured or manifest variable) used as a measure of a latent construct that cannot be measured directly. The researcher must specify a priori which indicators are associated with each latent construct.

Latent construct An operationalization of a construct in SEM. A latent construct cannot be measured directly but can be represented or measured by one or more measured variables (indicators).

Manifest variable An observed or measured value obtained from cases or observational units.

Measurement error The degree to which the variables we can measure do not perfectly describe the latent constructs of interest. Sources of measurement error can range from simple data entry errors to the definition of constructs that are not perfectly defined by any set of measured variables. All constructs have some measurement error, even with the best indicators. The researcher’s objective is to minimize the measurement error. SEM can take measurement error into account in order to provide more accurate estimates of the relationships between constructs.

Measurement model An SEM that specifies the indicators for each construct and enables assessment of construct validity.

Operationalizing a construct The key process in the measurement model involving determination of the measured variables that will represent a construct and the way in which they will be measured.

Path analysis General term for an approach that employs simple bivariate correlations to estimate relationships in an SEM model. Path analysis seeks to determine the strength of the paths shown in the path diagrams.

Reliability The measure of the degree to which a set of indicators of a latent construct is internally consistent in their measurements. The indicators of highly reliable constructs are highly interrelated, indicating that they all measure the same concept. Individual item reliability can be computed as 1.0 minus the measurement error. Note that high reliability does not guarantee that a construct is representing what it is supposed to represent. It is a necessary but not sufficient condition for validity.
Structural model Set of one or more dependence relationship linking the hypothesized model’s constructs. The structural model is most useful in representing the interrelationships of variables between constructs.

Structural relationship A dependence relationship (regression-type) specified between any two latent constructs. Structural relationships are represented with a single-headed arrow and suggest that one construct is dependent upon another. Exogenous constructs cannot be dependent on another construct. Endogenous constructs can be dependent on either exogenous or endogenous constructs.

2 What is Structural Equation Modeling?

Structural equation modeling is a general and flexible multivariate analysis technique that includes several other techniques as special cases. It can be viewed as melding of confirmatory factor analysis and linear regression. Indeed, factor analysis, principal components and partial least squares regressions, and linear regression all have elements in common with structural equation modeling and can even be seen as different kinds of structural equation models.

Major applications of structural equation modeling include:

1. causal modeling, or path analysis, which tests hypotheses of causation with a system of linear equations. Causal models can involve either manifest variables, latent variables, or both.

2. confirmatory factor analysis, an extension of factor analysis in which a priori hypotheses about the correlation structure of the factors are tested.

3. specified regression models, an extension of linear regression analysis in which regression weights may be constrained to be equal to each other, or to specified numerical values.

4. covariance structure models, which hypothesize that a covariance matrix has a particular form. For example, you can test the hypothesis that a set of variables all have equal variances with this procedure.

SEM can be used to model interactions, nonlinearities, correlated independent variables, measurement error, correlated error terms, multiple latent independent variables each measured by multiple indicators, and one or more latent dependent variables each also with multiple indicators. SEM is often a more powerful alternative to a host of other types of statistical models such as multiple regression, path analysis, factor analysis, time series analysis, and analysis of covariance. SEM arises directly out of the theory of the general linear model (GLM).

The advantages of SEM compared to multiple regression include more flexible assumptions, such as: allowing interpretation with multicollinearity, reduction of measurement error by having multiple indicators per latent variable, graphical model display, the ability to test inter-related models rather than individual model coefficients separately, the ability to test models with multiple dependent variables, the ability to model mediating variables, the ability to model error terms, and the ability to fit messy data such as time series with autocorrelated error, non-normal data, incomplete data.
2.1 The Basic Idea Behind Structural Modeling

As you recall from your elementary statistics courses, if every observation in a sample, with mean \( \bar{X} \) and standard deviation \( s \), is multiplied by some constant \( k \), then the new sample mean is \( k\bar{X} \). Similarly, the new standard deviation is \( |k|s \).

For example, suppose you have a sample consisting of \{1, 2, 3\}. The sample mean is 2 and the standard deviation is 1. If we took the sample observations and multiplied them by -5, then the sample mean would now be equal to -10, and the standard deviation would now be equal to 5, with the variance equal to 25.

In general, if a set of numbers \( X \) is related to another set of numbers \( Y \) by the equation \( Y = kX \), then the variance of \( Y \) must be \( k^2 \) times that of \( X \), so we can test the hypothesis that \( Y \) and \( X \) are related by the equation \( Y = kX \) indirectly by comparing the variances of the \( Y \) and \( X \) variables.

This idea generalizes, in various ways, to several variables inter-related by a group of linear equations. The rules become more complex, the calculations more difficult, but the basic message remains the same – you can test whether variables are interrelated through a set of linear relationships by examining the variances and covariances of the variables.

Recall that when we were considering the distribution of \( Y \), the response variable in a linear regression, we derived the variance thusly:

\[
Var \{Y|X\} = Var \{\beta_0 + \beta_1 X + \epsilon\} = Var \{\beta_0\} + Var \{\beta_1 X\} + Var \{\epsilon\} = Var \{\epsilon\} = \sigma^2
\]

The reason this derivation is so simple is that we are conditioning on a value of \( X \) associated with the value of \( Y \), hence \( Var \{\beta_1 X\} = 0 \). However, if \( X \) is a random variable independent of \( \epsilon \) with no conditioning on the observed value, then we need:

\[
Var \{Y\} = Var \{\beta_0 + \beta_1 X + \epsilon\} = Var \{\beta_0\} + \beta_1^2 Var \{X\} + Var \{\epsilon\} = \beta_1^2 \sigma_X^2 + \sigma^2
\]

The equation becomes more complicated as more and more predictor variables are added, especially if the predictor variables are correlated. However, this is the basic idea behind structural equation modeling.

In structural equation modeling, as in factor analysis and principal components analysis, we are concerned with the covariance matrix. The covariance of two random variables, \( X \) and \( Y \) is given by

\[
Cov \{X,Y\} = Cov \{Y,X\} = E \{XY\} + E \{X\} E \{Y\}
\]

The correlation of two random variables, \( X \) and \( Y \) is given by

\[
Corr (X, Y) = \frac{Cov (X, Y)}{\sqrt{Var (X) Var (Y)}}
\]
When $X$ and $Y$ are linear functions of each other, this value can be derived easily. If the linear relationship is exact, $\text{Corr}(X, Y) = 1$. As the number of random variables becomes large and the number of interrelationships considered increases, deriving the covariances can become quite difficult. Recall in factor analysis we were interested in the correlation between a factor and the variables. In SEM, we are concerned with the correlations between factors and variables as well as between the different factors themselves. The resulting system of linear equations can become mind-bogglingly complicated. Structural equation modeling was not practical as a data analysis tool until the advent of very fast micro-processors in the 1980s.

However, the bottom line is simple: if we make assumptions about the linear relationships between variables and between factors, we are explicitly making assumptions about their variance and covariance. These assumptions can be tested and this test forms the basis for structural equation modeling.

The basic procedure for SEM is as follows:

1. State the theory as to how the variables are inter-related and define the latent constructs (factors), often with the use of a path diagram.
2. Derive the expected variances and covariances of the constructs predicted by theory.
3. Test whether the variances and covariances fit the expected model.
4. Report the results of the hypothesis test, the parameter estimates and standard errors for the numerical coefficients in the linear equations.
5. Finally, assess over-all model fit.

Keep these 5 steps in the back of your mind as we work through the application of SEM. Although the mathematical and computational effort required to perform structural equation modeling is extremely complicated, the basic logic is embodied in these 5 steps.

A small caveat
Also, remember that a structural model, like any model, is only an approximation of Reality. The sensible question is not so much, “Is the model true?” but rather, “Is the model useful?” We must remember that just because a model fits the data well does not mean that the model is a good one. This is called the “fallacy of affirming the consequent”. If the model is a good one, then the data will fit the model. However, the data fitting the model does not imply that the model is good. The model fitting the data does not necessarily imply the model is good. There may be another model that fits the data equally well, or better.

2.2 Incorporating Latent Variables
One of the unique advantages of structural equation modeling is the ability to incorporate latent variables or latent constructs (these terms are used interchangeably). It is often the case that we are interested in somewhat abstract concepts and causalities that we cannot measure directly. The concept of intelligence is perhaps the most well-known. It has become accepted by researchers in psychological fields that some meaningful quantity related to our abstract concept of intelligence can be measured by the responses of the individual on a standardized test. Structural equation modeling
allows us to model other abstract concepts such as stability (say in an ecosystem), happiness, job satisfaction and others, through the latent construct approach. We have seen this basic idea when we “named” the factors in factor analysis. As scientists, our first job in structural equation modeling is to define our latent constructs. Eventually, we will create a list of variables that we can actually measure, called indicators or manifest variables, interchangeably that we believe together measure the latent constructs of interest. We will then incorporate the sampling of these variables into our experimental design.

2.2.1 Improving Statistical Estimation

In addition to being able to model abstract concepts, the above process also improves statistical estimation by providing an estimate of reliability and directly accounting for measurement error. In the various regression methods that we considered, it was implicit that the linear coefficients were estimates of the true coefficient. We can represent this concept by

$$\beta_{\text{obs}} = \beta_{\text{true}} \times \rho_x$$

Here, $\rho$ represents the reliability and is always between 0 and 1, implying that the observed correlation and the resulting coefficient will always underestimate the true strength of the relationship. Reliability is a measure of the degree to which a set of manifest variables is internally consistent with respect to representing a given latent construct. This measure is based on the how highly interrelated are the manifest variables in the set.

The concept of reliability corresponds directly to the concept of measurement error. In structural equation modeling this concept is extended beyond what we consider in other statistical techniques. In normal regression, the error term represented all unexplained variability. We know that whatever the quantity we are measuring, we cannot measure it perfectly. Sometimes there will be human error, sometimes our cases will lie, we will have instrumental measurement error, and there are other sources of variation that we are not including in the model, etc. Here, the concept of measurement error also includes error due to imperfect construction of the abstract concept we are interested in, or in other words, in the construction of the latent construct. If the manifest variables are all highly related, then we have high reliability and low measurement error. However, high reliability does not necessarily imply high validity. We will discuss the concept of validity in the next chapter.

2.2.2 Exogenous versus Endogenous Latent Constructs

In SEM, we have multivariate analogs to response and predictor variables. Because in SEM we are usually dealing with latent constructs, we have a different vocabulary.

*Exogenous constructs* are the multivariate analog to the predictor variable. The exo (exo=out) root means that we never try to predict or explain exogenous variables from other variables or constructs in the model. Hence, exogenous constructs are affected only by factors outside the model. In a path diagram, these constructs have arrows pointing away from them, never towards them from any other variable or construct. (The arrows linking the construct to its manifest variables point from the construct to the variable). An example is intelligence, which could be
measured by a test with 100 questions. The respondents answers to the 100 questions are the 100 manifest variables that define the exogenous construct, intelligence.

Endogenous constructs are the multivariate analog to response variables. These constructs are theoretically determined by other variables or constructs in the model. In path diagrams, these constructs have arrows pointing towards them, but can also be inter-dependent with other endogenous constructs or variables, so they can have double-headed arrows pointing away and towards them.

Continuing with the example of intelligence, we might be interesting in how intelligence and job satisfaction relate to marital satisfaction. Intelligence would be measured by our 100 question IQ test. Job satisfaction and marital satisfaction might be measured by the respondents answers to different tests (instruments) of 10 questions each. The 10 answers on the job satisfaction instrument are 10 manifest variables that define the exogenous construct, job satisfaction. Likewise, the 10 answers to questions on the marital satisfaction instrument are the 10 manifest variables that define the exogenous construct, marital satisfaction.

The exogenous constructs, intelligence and job satisfaction, would then be used to predict the endogenous construct marital satisfaction.

2.2.3 The Measurement Model and the Structural Model

In SEM terminology, the complete model is composed of two layers: the measurement model and the structural model. The measurement model represents how the measured variables are related to one another and to which constructs, while the structural model represents how the constructs are related to one another. These models must both be conceptually (and usually graphically) specified before the analysis, and preferably before the data are even gathered.

2.3 Structural Equation Modeling and the Path Diagram

A popular approach to structural equation models is to express them as path diagrams. This helps investigators both in developing and comprehending their own models and in presenting their theory and the model-fitting results to their audiences. We will be drawing these on the board quite often and it is something you should practice on your own.

Path diagrams play a fundamental role in SEM. Path diagrams are flowcharts with specialized symbols, such as lines, arrows, boxes and circles, to indicate variable characteristics and relationships.

Consider the classic linear regression equation. We’ll drop the intercept term for simplicity:

\[ Y = \beta_1 X + \epsilon \]

We can represent this equation in a path diagram as follows:
All variables in the equation system (theory) are placed in the diagram, either in boxes or ovals. Each equation is represented on the diagram as follows:

- Exogenous constructs have arrows pointing to endogenous variable(s).
- Variables for exogenous constructs are represented with $Y$ and those for endogenous with $X$.
- The weighting coefficient is placed above the arrow.
- Constructs are thought to cause manifest variables and so arrows point from constructs to manifest variables.
- Manifest variables are placed in rectangles in the path diagram.
- Latent constructs are placed in an oval or circle. For example, the variable $\epsilon$ in the above diagram can be thought of as a linear regression residual when $Y$ is predicted from $X$. Such a residual is not observed directly, but calculated from $Y$ and $X$, so we treat it as a latent variable and place it in an oval.
- The absence of an arrow means that there is no relationship between the constructs according to the theory.
- Straight one-headed arrows depict causational relationships and curved two-headed arrows depict correlational relationships.

The example discussed above is an extremely simple one. Generally, we are interested in testing models that are much more complicated than these. As the equation system becomes increasingly complicated, so does the covariance structure.

What causal modeling allows us to do is examine the extent to which data fail to agree with one reasonably viable consequence of a model of causality. If the linear equations system derived from the path diagram does fit the data well, it is evidence in favor of the causal model, though of course, it is not proof that the model is true.

Now let’s try diagramming the intelligence and satisfaction model.

Path Diagram for Intelligence and Job and Marital Satisfaction Model

2.3.1 Specification of the Theory

SEM should never be attempted without a strong theoretical basis for the specification of the causational and correlational relationships. While this is true to some extent for all statistical
analyses, it is particularly true for SEM because it is considered a confirmatory analysis. If we
don’t have any theory, then there’s nothing to confirm. If our theory is weak, then the addage
“Garbage in, garbage out” is relevant.

Causation is perhaps the strongest sort of theoretical inference. You have probably had it
drilled into your heads that “correlation is not causation” and that we cannot test causation with a
regression model from an observational study. However, how many of you would buy the argument
that it’s unclear that cigarettes cause lung cancer because maybe lung cancer causes cigarette
smoking? Why not?

How can we test a theory involving causation? In controlled experiments, such as drug trials, we
can infer causation. This is partly because controlling experimental conditions along with random
assignment and replication should eliminate any confounding factors and spurious relationships.

A confounding factor occurs when we cannot tease apart the effect of two variables on a third.
For example, if we have two tanks and apply one treatment to one tank and another treatment to
the other, treatment and tank are confounding one another. Any effect we observe could be either
from the treatment or from the tank.

A spurious relationship is when we observe a strong correlation between two variables that
is actually due to the fact that both variables are strongly correlated with a third, unmeasured
variable.

For example, if we assign the first 30 fish we caught to the first treatment and the next 30 to
the second treatment, then any relationship between treatment and our outcome could be spurious
because it may be due to the ease with which we caught the fish (maybe we are using a net and
we will catch the large (or dumb) fish first. Note that treatment and catch order are confounded.

If we adequately control, replicate and randomly assign, then any effect we observe must (almost
certainly) be due to the treatment, except for the small probability of a type I error.

Designs that involve random sampling from a population, observational studies, usually cannot
be used to test causational hypotheses. However, the evidence against cigarettes is mostly from
observational studies. No one has randomly assigned a large number of replicates (humans) to a
‘cigarette’ treatment and to a ‘no cigarette’ treatment. Why are we convinced that smoking causes
cancer and not the other way around?

Requirements to Establish Causation

1. Covariation: Causality implies directly that a change in the cause will bring about a change
in the effect. Hence, covariation between two variables is necessary but not sufficient to
establish a causational relationship between them. SEM tests for significant covariation be-
tween constructs and statistically significant estimated paths in the structural model provide
 evidence of covariation.

2. The effect must follow the cause in a temporal sequence: If the probability of smoking
increased after the presence of lung cancer, we might start to believe that smoking is caused
by lung cancer and not the other way around. SEM cannot provide evidence of a temporal
sequence in an observational study unless it is longitudinal. A controlled experiment will also
provide evidence for the sequence.
3. **Nonspurious Covariance**: This is tested mainly by looking to see if significant covariation remains in the presence of other possible causes. This is easy to test when there is no collinearity in the predictors. The addition or deletion of other variables will not significantly affect the strength of the covariance. However, most multivariate data will have some degree of multicollinearity. If the constructs are well-defined, the covariation between constructs should still remain reasonably stable with the addition of others. Hence, causational relationships can still be tested so long as the constructs are well-defined.

4. **Theoretical Support**: If there is strong scientific theory to support a causational relationship, then strong covariation and resilience against tests for spurious covariation are highly convincing that the relationship exists. Without this third requirement, all other arguments are weak.

3 A Simple Example of SEM

We’ll work through the example in the book and then study a more complicated example with real data.

3.1 **The Research Question**

For this example we wish to predict customer share (a measure of how much a customer spends on the brand or store versus competing brands or stores) and customer commitment (a measure of how much a customer feels emotionally attached to and is willing to sacrifice to obtain/shop at the brand/store). We would like to model how customer perceptions of three key strategy elements – price, service and atmosphere – determine customer acceptance of the store, measured by customer share and customer commitment. Retailers believe that favorable impressions of price, service and atmosphere encourage customers to return and spend more money (increases customer share) and that this leads to increased customer commitment. Because of the multi-level problem, this model is not tractable with multiple regression.

**The Conceptual Model**

1. Favorable price impressions increase customer share.
2. Favorable service impressions increase customer share.
3. Favorable atmosphere impressions increase customer share.
4. Increased customer share increases customer commitment.

<table>
<thead>
<tr>
<th>Exogenous Constructs</th>
<th>Endogenous Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Customer Share</td>
</tr>
<tr>
<td>Service</td>
<td>Customer Commitment</td>
</tr>
<tr>
<td>Atmosphere</td>
<td></td>
</tr>
</tbody>
</table>

Once the constructs have been specified, the next step is to specify the relationships between the constructs. We have already specified the causal relationships in the conceptual model, but
have not specified any correlations between the exogenous constructs. We would expect separate elements of retail strategy to be associated as part of consistent planning and execution on the part of the retailer, so we will allow for correlations between the exogenous constructs.

Path Diagram for the Customer Commitment Model

3.2 Estimation and Assessment

Once the theory is established and the data collected the SEM analysis is run through a computer program to fit the model and estimate the parameters. The parameters are the variances and covariances and the coefficients, the SEM analogs to the coefficients in linear regression.

3.2.1 The Observed Covariance Matrix

In spite of the title of this section, we can actually use either the covariance matrix or the correlation matrix. There are pros and cons, mostly that arise out of old-style programming techniques that required the user to input the covariance/correlation matrix. Nowadays, most programs can work with the raw data and so both the covariance and correlation matrices can be easily calculated. Keep in mind that the correlation matrix is merely a scaled covariance matrix, where the covariances of the normalized variables are calculated. We will discuss this more in the next chapter.

A covariance matrix is symmetric—the values above the diagonal are equal to the values below the diagonal. This is because \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \). Often, just the lower triangle of the matrix is shown.

A general covariance matrix \( \mathbf{S} \) for our example, with constructs price, \( P \); service, \( S \); atmosphere, \( A \), customer share, \( CS \) and customer commitment, \( CC \); looks like this

\[
\mathbf{S} = \begin{bmatrix}
\text{var}(P) & - & - & - & - \\
\text{cov}(P, S) & \text{var}(S) & - & - & - \\
\text{cov}(P, A) & \text{cov}(S, A) & \text{var}(A) & - & - \\
\text{cov}(P, CS) & \text{cov}(S, CS) & \text{cov}(A, CS) & \text{var}(CS) & - \\
\text{cov}(P, CC) & \text{cov}(S, CC) & \text{cov}(A, CC) & \text{cov}(CS, CC) & \text{var}(CC)
\end{bmatrix}
\]
Suppose the observed values of $S$ are as below:

$$S = \begin{bmatrix}
\text{var}(P) & - & - & - & - \\
0.20 & \text{var}(S) & - & - & - \\
0.20 & 0.15 & \text{var}(A) & - & - \\
0.20 & 0.30 & 0.50 & \text{var}(CS) & - \\
-.05 & 0.25 & 0.40 & 0.50 & \text{var}(CC)
\end{bmatrix}$$

If the data were standardized, the off-diagonal elements would be correlations and the diagonal elements would all be equal to 1.

Say the parameter estimation returned the following:

**Path Diagram for the Estimated SEM for Customer Commitment**

The path coefficients represent the strength of the relationship between the constructs. Recall that in factor analysis, the loadings or coefficients are the correlations between the factor and the variable.

Considering first the model predicting customer share, we see that atmosphere has the largest influence, followed by service, while price has the smallest influence. Next, we see that customer share has a substantial impact on customer commitment, with a coefficient of 0.5.

We can go a step further and obtain a predicted value for either endogenous variable. From the path diagram, we see that

$$\hat{y}_{cs} = 0.065P + 0.219S + 0.454A$$
$$\hat{y}_{cc} = .50CS$$
$$\Rightarrow \hat{y}_{cc} = 0.5 [0.065P + 0.219S + 0.454A]$$

Note that the variables in these equations are constructs and so the units are not necessarily those of the original measured variables.
3.2.2 Assessing Model Fit

Assessing model fit boils down to comparing the observed covariance matrix to the estimated covariance matrix based on the model. As we discussed earlier, a covariance structure can be derived from the model. If the model fits the data, there will be consistency in the model estimated covariance structure and the observed. Let’s see how this is done. Here, “consistent” means “the same plus or minus sampling error”. We will discuss the formal goodness-of-fit tests later in the chapter.

Let’s take one covariance and examine how it is estimated according to the model. The estimated covariance is the sum of the direct path covariances (loadings) and the indirect path covariances (loadings). For $\text{Cov}(S, CS)$ we have:

Estimated $\text{Cov}(S, CS)$

**direct path:**

- service $\rightarrow$ customer share $= 0.219$

**indirect paths:**

- service $\rightarrow$ price $\rightarrow$ customer share $= 0.200 \times 0.065 = 0.013$
- service $\rightarrow$ atmosphere $\rightarrow$ customer share $= 0.150 \times 0.454 = 0.068$

**Total:** direct + indirect:

$0.219 + 0.013 + 0.068 = 0.300$

Let’s try another one. The estimated covariance for atmosphere and customer share is:

Estimated $\text{Cov}(A, CS)$

**direct path:**

**indirect paths:**

**Total:** direct + indirect:

By a procedure such as this the estimated covariance matrix can be constructed, usually, of course, by the software program and not the investigator.

For this example, the estimated covariance matrix, $\Sigma_k$ is:

$\Sigma_k = \begin{bmatrix}
\text{var}(P) & - & - & - & - \\
0.20 & \text{var}(S) & - & - & - \\
0.20 & 0.15 & \text{var}(A) & - & - \\
0.20 & 0.30 & 0.50 & \text{var}(CS) & - \\
0.10 & 0.15 & 0.25 & 0.50 & \text{var}(CC)
\end{bmatrix}$

As we see, some of the estimated covariances are equal to the observed and some of quite different. The difference between the observed and estimated (predicted) covariances is called the
residual. Note that in this case the residual is still the difference between the observed and predicted, as we saw in linear regression, only here it is the difference between observed and predicted covariances and not observed versus predicted individual observations or means.

3.3 Model Development Strategies

When attempting to test scientific hypotheses with high dimensional data, often the optimal approach is a blend of confirmatory (testing a specific *a priori* hypothesis) and exploratory (letting the data suggest or influence the specification of the model) approaches. The strategies we will examine in this section include confirmatory modeling strategy, competing model strategy, and model development strategy.

3.3.1 Confirmatory Model Strategy

In this strategy the investigator specifies a single model (a specified set of relationships) and SEM is used to test how well the data fit the model. The investigator would report the results of the goodness-of-fit tests and that would be the end of the study. If the data do not fit the model, then seeking a better model would be part of an on-going research effort or debate within the field. If the data do fit the model, the investigator cannot state that the model is true or that the model is the best model for the problem. He or she can only state the data are consistent with the model.

For all practical purposes, this approach is almost never used, except when scientists are arguing and seeking to disprove each others models. Perhaps a more fruitful approach for individual scientists is to compare several competing models as part of a single study.

3.3.2 Competing Models Strategy

As the name implies, scientists using this strategy would propose several models that they find scientifically interesting and plausible and test them all using SEM. This approach is more powerful than the confirmatory approach in that a definitive statement can be made about which of the models tested is the best one. Of course, we still can’t say that the chosen model is the best of all possible models.

How are the competing models chosen? One obvious choice of a competing model would be a model that has been reported in the literature but that one feels can be improved by additional or contrary theory. That is, a scientist may wish to show that his idea or discovery can improve prediction or mechanistic understanding of the phenomenon better than existing models. Also, an individual scientist may have several competing ideas or hypotheses as to the dynamics of a system. These ideas will generate competing models to be tested.

The downside of the competing model strategy is that it is possible, if not probable, that the competing models will fit the data equally well. It has been shown that for any structural equation model, at least one other model exists that has the same number of parameters (though different specified relationships) that fits the data equally well. These models are called equivalent models.
3.3.3 Model Development Strategy

The most common approach reported in the literature when SEM is used is a model development strategy. In this approach, a basic framework for the model is specified, but the measurement and structural models are modified in response to goodness-of-fit tests until a good fit is found. Here the results of iterative SEM analyses provide insight that lead to better models.

The problem with this strategy is that too much respecification of the model in response to goodness-of-fit tests can lead to over-fitting and a loss of generalizability. We will explore this strategy in the following chapters.

4 The Five Stages in Structural Equation Modeling

Let’s review all we’ve discussed so far and go into a little more detail by outlining the steps of a structural equation modeling analysis.

The five stages are as follows:

1. Defining individual constructs and developing the overall measurement model
2. Designing a study to produce empirical results
3. Assessing the measurement model validity
4. Specifying the structural model
5. Assessing structural model validity

We will work through these stages again to reinforce the concepts we’ve been discussing. We’ll continue to use the customer commitment model as well as generalized models.

4.1 Defining Individual Constructs and Developing and Specifying the Measurement Model

A well specified measurement model is a prerequisite for useful results from an SEM. Any resulting hypothesis tests be no more useful than the theory behind the building of the constructs. An investigator Operationalizes the construct once it has been defined, by identifying which measured variables define it. In survey research, this usually involves several questions that apply to the concept upon which the construct is based. If prior research has already operationalized the construct, then the variables associated with it in the literature should be used, unless there is justification to change the operationalization.

The measurement model can be specified by a system of equations, but it is easier to represent the model with by our now familiar path diagram. We will add a little more complexity to represent complete model.
4.2 Designing a Study to Produce Empirical Results

Once the basic model is specified, the researcher is in an excellent position to develop a strong experimental design. As with other multivariate techniques, there are some rules-of-thumb for experimental design of an SEM analysis.

4.2.1 Covariance versus Correlation

Nowadays the issues are not particularly important. It’s easy to input the raw data and have the program calculate both the covariance and the correlation matrices. If the researcher does not have the raw data, it’s easy to input the covariance matrix and ask for the program to output the correlation matrix. The advantage of the correlation matrix are that the parameter estimates are standardized and thus are not scale dependent. The advantage of the covariance matrix is that using the correlation matrix as input can lead to error in the computation of the standard errors. Hence, it’s recommended to use the covariance matrix if you don’t have the raw data and ask for correlation matrix as output.

4.2.2 Missing Data

Please review this section in your notes and text.

4.2.3 Sample Size

Determining optimal sample size is more complicated for SEM, but the issues are similar as to those in other multivariate techniques.

Multivariate Distribution As the distribution of the data deviates more from normality, then the ratio of respondents to parameters increases. A rule-of-thumb 15 respondents for every parameter. As we have seen, the number of parameters in an SEM can be huge. Striving for
parsimony is always a good idea, but is even more important in SEM. It becomes even more so with irregular distributions.

**Estimation Technique** By far the most common approach is maximum likelihood estimation (MLE). Maximum likelihood estimation is asymptotically unbiased and has been shown to work for even small sample sizes (such as 50). However, the recommendation is 100 to 150 observations are recommended to produce stable solutions.

It should also be noted that as the sample size becomes large, almost any difference between the observed and estimated covariance matrices will become significant and the model will fail the goodness-of-fit test. So, it’s more important here not to have more observations than are necessary. This isn’t likely to be the case very often because SEM requires such a large amount of data.

**Model Complexity** Simpler models can be tested with smaller samples. This includes not only complexity due to a large number of measured variables, but also due to models with a large number of constructs, constructs with fewer than 3 indicators, and models with several groups. All of these situations require larger sample sizes. Inadequate sample sizes lead to unstable and highly variable parameters estimates and sometimes to problems with convergence of the algorithm.

**Missing Data** Missing data complicate the fitting of SEMs, as they do with most techniques.

**Average Error Variance of Indicators** If there is a good deal of variance not explained by the model we often observe problems with convergence and model stability. Small communalities indicate that the associated constructs do not adequately explain the variation in the indicators. Models containing multiple constructs with communalities less than 0.5 require larger sample sizes.

**Summary of Sample Size Guidelines**

- SEMs with 5 or fewer constructs that each have at least three indicators with high communalities require a sample size of only 100-150.
- If any of the communalities are between 0.45-0.50, or the model contains constructs with fewer than three indicators, then the required sample size is around 200.
- If any of the communalities are less than 0.45, or the model includes more than one construct with fewer than 3 indicators, then the minimum sample size is around 300 or more.
- When the model contains more than 6 constructs where some of these have fewer than 3 indicators and multiple small communalities, then the sample size required for stability may exceed 500.

**4.3 Assessing Measurement Model Validity**

There are various ways to check the validity of the measurement model. We consider several of them here. Measurement model validity depends on a combination of goodness-of-fit and specific evidence of *construct validity*. 


4.3.1 Goodness-of-fit

Goodness-of-fit (GOF) for a measurement model in SEM is defined as how well the model reproduces the observed covariance matrix of the indicators, i.e. how close the observed covariance matrix is to the estimated. A variety of GOF statistics have been developed, each striving to improve on the other or to focus on different aspects of the model’s ability to explain the data. There are three specified categories of GOF statistics: absolute measures, incremental measures, and parsimony fit measures. SAS outputs a very long list of GOF statistics automatically. We will consider only a few of them here.

The Basics of Goodness-of-fit Statistics Most goodness-of-fit statistics are based, one way or another on the $\chi^2$. The main advantage of the $\chi^2$ statistic it is the only GOF statistic whose distribution is known and therefore the only one that can be used to conduct formal hypothesis testing, rather than merely using rough guidelines to make a decision.

The $\chi^2$ GOF The $\chi^2$ statistic is based on the difference in the covariance matrices. Most SEM estimation procedures, such as maximum likelihood, seek to minimize this difference. A $\chi^2$ test tests the significance of this difference. The test statistic is given by:

$$\chi^2 = (N - 1) (S - \Sigma_k)$$

where $N$ is the total sample size. The distribution of this statistic depends not only on sample size but on the number of parameters, through the degrees of freedom.

Degrees of Freedom The degrees of freedom of a statistic represent, in a vague sense, the amount of information available to estimate model parameters. Usually, the degrees of freedom depend on sample size. The $\chi^2$ statistic in the SEM goodness-of-fit test has degrees of freedom that depend only on the number of unique elements in the covariance matrix. The formula for the degrees of freedom is given by

$$df = \frac{1}{2} [p(p + 1)] - k$$

where $p$ is the number of observed variables and $k$ is the number of estimated (free) parameters. The first term in the difference is the number of covariance terms below the diagonal plus the number of variances on the diagonal. Notice that the degrees of freedom do not depend on sample size. This has important implications on the use of $\chi^2$ as a GOF statistic.

Statistical Significance of $\chi^2$ The null hypothesis in the $\chi^2$ test is that the two covariance matrices are equal. The larger the difference between the two matrices, the larger the $\chi^2$ statistic. Hence, large values of $\chi^2$ indicate lack of fit. SEM analysis packages, such as SAS, supply both the value of the statistic and the $p$-value associated with it. For this test, we want to see large $p$-values indicating no significant difference between our observed and model-estimated covariance matrices.

Absolute Fit Measures Absolute measures of fit compare how well a model fits the data, independently from any other model. Hence, they do not explicitly compare a given model to
another. Absolute measures of fit provide the most basic measure of how well a model fits the data.

\( \chi^2 \) statistic, revisited The most fundamental GOF statistic is the \( \chi^2 \) as discussed above. Theoretically, it is the same statistic used in categorical data analysis (contingency tables and cross-classification). However, in categorical data analysis the investigator’s theory is validated by large values of the \( \chi^2 \) statistic indicating that the data do not fit the null hypothesis of no difference, and hence strengthening the alternative hypothesis of a treatment effect or of multiple populations. In SEM models, the investigator’s theory is strengthened by small values of the \( \chi^2 \) indicating that the observed data do not differ significantly from what is predicted by the model. These two approaches are logically very different from one another.

The \( \chi^2 \) statistic has two mathematical properties that make it somewhat problematic as a GOF statistic. First, as you will notice, the statistic becomes larger (more significant) with sample size. This is the case with most statistics which is why power (the ability to detect deviations from the null hypothesis) increases with sample size. But for SEM GOF tests, even if the covariances matrices are quite close to one another, with large sample sizes the \( \chi^2 \) becomes more and more significant, falsely indicating a lack-of-fit for the model. Though not mathematically obvious, the probability that a \( \chi^2 \) will be large increases with the number of free parameters in the model as well. Since SEM models tend to be large, this can cause problems where it might not in other analyses (though perhaps it could be argued that there should be a penalty for a large number of parameters). In any case, the \( \chi^2 \) can produce apparent lack-of-fit when the model is actually performing well. For this reason, researchers have developed other GOF statistics, hoping to improve upon \( \chi^2 \).

Goodness-of-Fit Index (GFI) The GFI is an early attempt to create a GOF statistic that is insensitive to sample size. Though the sample size does not appear in the formula, there is still some influence of sample size due to the effect of \( N \) on sampling distributions. The GFI ranges from 0 to 1, with higher indicative of good fit. Let \( F_k \) be a measure of the variance explained by the model with \( k \) parameters and \( F_0 \) be a measure of the variance explained by the model where all the parameters are equal to 0, i.e., none of the constructs are related to one another. Then \( GFI \) is given by:

\[
GFI = 1 - \frac{F_k}{F_0}
\]

GFI measures the relative amount of the variance in \( S \) that is predicted by \( \Sigma_k \). \(^1\)

Root Means Square Residual (RMSR) and Standardized RMSR (SRMR) The RMSR is the square root of the mean of squared residuals of the covariance matrix: the average

\[
GFI = 1 - \frac{\text{tr} \left( (\Sigma_k^{-1} S - I)^2 \right)}{\text{tr} \left( (\Sigma_k^{-1} S)^2 \right)}
\]

\(^1\)For you linear algebra buffs:
of the squared residuals between the individual observed and estimated variances and covariances. The residuals are taken as the observed minus the predicted. The expected value of the residual under the null hypothesis (the data fit the model) is 0. When covariances are used the residuals are in the units of the original measurements, so it’s difficult to compare RMSRs across different models. The solution to this problem is the standardized RMSR, or the residuals between the observed and estimated correlation matrix. The problem with RMSR and SRMR is that their distributions are unknown and hence there are no established thresholds (we can’t find p-values). However, practical significance can be assessed by the researcher. Lower values of RMSR and SRMR indicate better fit and higher values indicate worse fit, leading some to dub these measures as badness-of-fit measures. Individual SRMRs enable the researcher to identify problematic manifest variables and problems with the measurement model.

Root Mean Square Error of Approximation (RMSEA) RMSEA is another measure of absolute fit that attempts to improve upon the $\chi^2$ statistic’s tendency to be more likely to reject models with large sample sizes. It also improves upon RMSR and SRMR because it has a known distribution (the non-central $\chi^2$). Thus, we can establish p-values and confidence intervals for our measure of fit.

As before, $N$ is the sample size and $df$ is the degrees of freedom. RMSEA is given by

$$RMSEA = \begin{cases} \sqrt{\frac{\chi^2 - df_k}{N - 1}} & : \text{when } \chi^2 - df_k > 0 \\ 0 & : \text{when } \chi^2 - df_k < 0 \end{cases}$$

As with RMSR, better models have lower RMSEAs. An empirical examination of several measures found that RMSEA was best suited for use in confirmatory or competing model strategies with large sample sizes (more than 500).

Actual Cross-Validation Index (CVI) The actual cross validation index is found by finding a matrix analog of the PRESS statistic using the computed covariance matrix derived from a model in one sample to predict the observed covariance matrix taken from a validation sample. This index requires splitting the data set into a training set and a test set, and so requires a large amount of data.

Expected Cross Validation index (ECVI) As with other cross-validation statistics, the ECVI is a measure of how well the model would fit data not used during the model-fitting procedure. The ECVI takes model complexity and sample size into account and is useful for comparing the performance of one model to another. It is linearly related to AIC. Smaller values of ECVI indicated better fit. As its name implies, it is the expected value of the actual cross validation index. It is given by

$$ECVI = F(S, \Sigma_k) + \frac{2p}{N - n_x - 1}$$

where $F(S, \Sigma_k)$ is the maximized likelihood function, $p$ is the number of parameters, and $n_x$ is the number of manifest variables.
Incremental Fit Indices Incremental fit indices differ from absolute fit indices in that they compare a given model to an alternative, baseline model, usually the null model (the one that assumes no correlation between the constructs).

Normed Fit Index (NFI) The normed fit index is one of the original incremental indices. It’s given by

\[ NFI = \frac{\chi^2_k/d_k - \chi^2_0/d_0}{\chi^2_0/d_0} \]

It ranges from 0 to 1, with 1 indicating a perfect fit.

Comparative Fit Index (CFI) The CFI is derived from the NFI with an aim to include model complexity in the measure of fit. The CFI is one of the most widely used indices because of its many desirable properties, one of which is its relative insensitivity to model complexity. The CFI is given by

\[ CFI = \max \left\{ \frac{N(\chi^2_k - df_k)}{\chi^2_0 - df_0}, 0 \right\} \]

Parsimony Fit Indices This set of indices provides information about the relative goodness-of-fit for competing models. They all take into account fit relative to complexity. Like \( R^2 \), some goodness-of-fit measures will automatically improve with increasing model complexity, making it difficult to compare models with different number of parameters. Parsimonious fit indices take this into account. The degrees of freedom form the basis for comparing models across levels of complexity, usually in the form of the parsimony ratio (PR). The PR is the ratio of the \( df \) of the model to the total \( df \) available.

Information Criteria (AIC, CAIC, SBC) These three criteria are based on the information matrix. The AIC we have seen before. As with linear regression and logistic regression, lower values imply better fit.

\[ AIC = \chi^2 - 2df \]
\[ CAIC = \chi^2 - (\ln(N) - 1)df \]
\[ SBC = \chi^2 - \ln(N)df \]

Parsimonious Goodness-of-Fit Index (PGFI) The PGFI makes an adjustment of the GFI to account for the number of parameters. As with other parsimony measures, the use of the PGFI is in comparing two models of interest. The model with the with the larger PGFI is preferable. The PGFI is given by

\[ PGFI = \frac{df_k}{df_0} GFI \]

Parsimonious Normed Fit Index (PNFI) The PNFI adjusts the normed fit index for parsimony. Like the PGFI relatively large values indicate the better model. It is given by

\[ PNFI = \frac{df_k \chi^2_k - \chi^2_k}{df_0 \chi^2_0} \]

21
The use of goodness-of-fit statistics has generated some discussion, similar to that of model selection using other methods. The aim of SEM is to test theory, not to conduct a fishing expedition for the best fit. As with other types of modeling, scientific research is best served by well-defined theory and the testing of well-characterized models, rather than any sort of shot-gun approach that relies on the mindless process of trial and error until a good fit is found.

4.4 Specifying the Structural Model

Specifying the structural model involves specifying the relationships between the constructs based on the proposed theoretical model. The measurement model involves how the manifest variables are related to each other and any correlations between exogenous constructs. The structural model involves how the endogenous constructs are related to the exogenous constructs. The arrows connecting the constructs in the path diagram can be written out as formal hypotheses.

Let’s return to the customer commitment model, both to show the set of hypotheses to be tested and the path diagram for complete model, including the measurement and structural relationships.

Hypotheses for the Complete Model

- $H_1$: Consumer’s price perceptions are positively correlated with customer share.
- $H_2$: Consumer’s service perceptions are positively correlated with customer share.
- $H_3$: Consumer’s atmosphere perceptions are positively correlated with customer share.
- $H_4$: Customer share is positively correlated with customer commitment.

Note that stating relationships explicitly entails some relationships that are implicit. That is, in our diagram, the relationships between constructs that are not linked by arrows have zero correlation. We are specifying this when we show no arrows connecting price and customer commitment (or service or atmosphere and customer commitment). In this sense, the structural model is a more constrained model than a model leaving all parameters free. This is how our specific theory is specified and subsequently tested.

The final path diagram showing all relationships between constructs and their manifest variables and with each other can be drawn as shown below. The measurement model is shown with hatched lines and the structural model is shown with solid lines.

Once both the measurement model and the structural model have been specified, the overall theory is complete and ready for formal testing.
4.5 Assessing the Structural Model Validity

The final stage is assessing the overall fit and validity of the structural model. This stage is similar to stage 3. We are still using the observed and estimated covariance matrices to assess model fit. However, the estimated covariance matrix now includes the influence of the fixed parameters (the zero correlations) that define the structural model. Recall that the measurement model assumes that all constructs are correlated to one another. That is, there are no fixed parameters. In the structural model, some of the correlations are set to zero by the researcher, hence the structural model has a lower number of parameters to be estimated. Hence, the degrees of freedom are larger for the structural model than for the measurement model, that is, the structural model is more highly constrained.

The effect of this on the $\chi^2$ statistic is that the measurement model will always fit at least slightly better than the structural model, i.e., the $\chi^2$ statistic will be smaller. This is much like the error sum of squares in a linear regression, which will always be smaller with more parameters. The test of the validity of the structural model is to see whether the improved fit of the measurement model is significantly better than we would expect under random sample variation.

A standard approach to assessing the structural model validity is to examine the $\chi^2$ statistic, one other absolute fit index, one incremental index, and one parsimony index. These indices should be compared to those for the measurement model as well as for any competing structural model.

**Competitive Fit Assessments** Generally we will be comparing the structural model to some other model, either the measurement model or another complete structural model. We compare models by examining the change in the $\chi^2$ value (or any of the incremental or parsimony fit indices) as the model becomes more constrained (fewer estimated parameters).
Comparing Nested Models

The basic $\chi^2$ test is given by the follow equation:

$$
\Delta \chi^2_{\Delta df} = \chi^2_{df_r} - \chi^2_{df_f}
$$

where

$$
\Delta df = df_r - df_f
$$

and where the full model is the model with more free parameters and the reduced model is the one with fewer. If we are comparing the structural model to the measurement model, the measurement model is the full model. Now, because the difference between two $\chi^2$s is also distributed $\chi^2$, $\Delta \chi^2_{\Delta df}$ can be examined and a p-value found for the test of no difference. Again, the larger $\Delta \chi^2_{\Delta df}$, the stronger the support for the larger model.

Would a large $\Delta \chi^2_{\Delta df}$ provide support for or against the structural model when comparing it to the measurement model?

What would $\Delta df$ be equal to if we were comparing the measurement model for the customer commitment model to the structural model?

What would $\Delta df$ be equal to if we were comparing the customer commitment structural model with an additional path from the price construct to customer commitment?

Comparison to Other Competing Models

SAS outputs the AIC which is a model selection criterion, unlike most, that can be used to compare models that are not nested.

Equivalent Models

Models with different specified paths may produce nearly identical covariance matrices and goodness-of-fit statistics. The way to compare models with equivalent fit is to use our scientific understanding of the proposed model. Complicated models can have a large number of equivalent models, but many of them will not make any sense.

Testing Structural Relationships

The final step after model fit is assessed is to examine the parameter estimates to see if they support the structural model. There are three basic requirements

1. The hypothesized parameters and statistically significant and in the predicted direction.
2. The parameter estimates are non-trivial from a scientific perspective, i.e., the effect size is large enough to be scientifically interesting. Examining the standardized loadings is helpful here.
3. The researcher will also want to examine the amount of variation in the endogenous constructs explained by the model.

We will examine each point in the process as we work through an example in more detail in the next chapter.
5 Appendix

  http://www2.chass.ncsu.edu/garson/pa765/structur.htm


  http://www.statsoft.com/textbook/stsepath.html