Structural Equation Panel Models I: Introduction to SEM

PS 2701-2019

Longitudinal Analysis

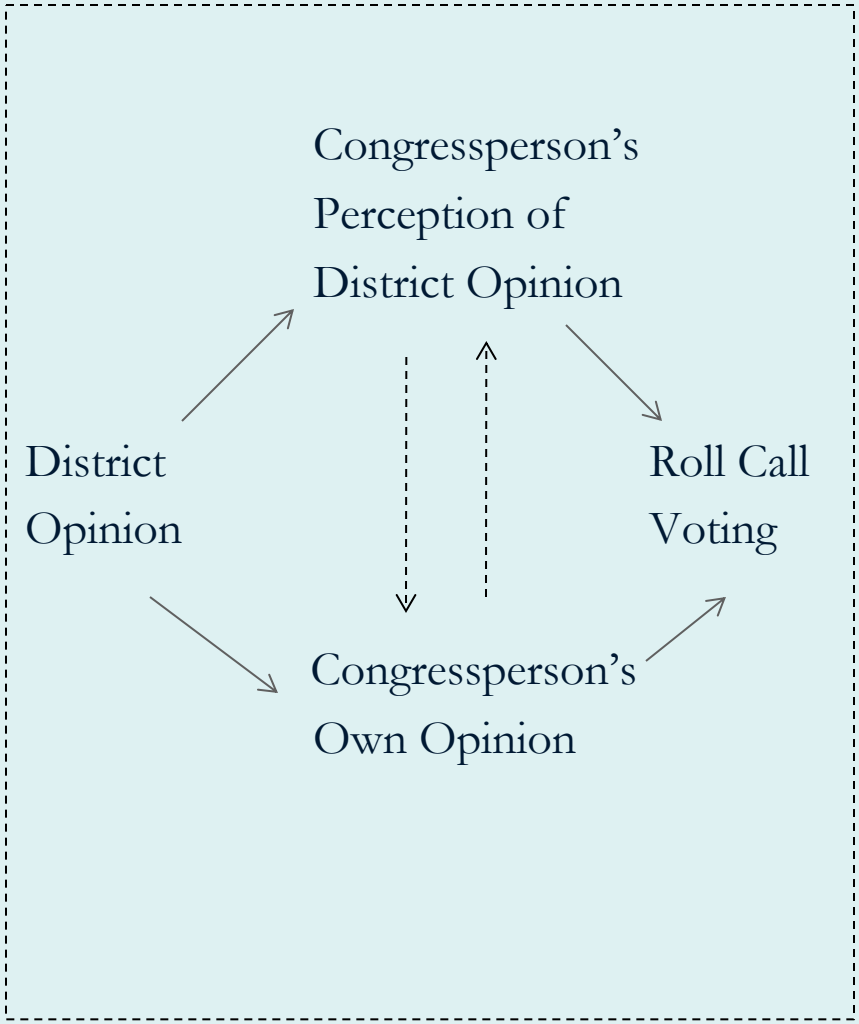
Week 5

Professor Steven Finkel

SEM: A General Analytic Framework

- Also referred to as "covariance structure analysis "(CSA)
- Method for analyzing *systems of equations*, how a series of variables may relate together or influence one another, either in a unidirectional causal sense or in a more complex models of "reciprocal" causality
- Present a system of equations that represents the causal linkages between variables, so that the model shows all ways that the variables are related. By estimating the parameters in the model *as a whole*, as opposed to equation by equation, we gain information to test more complex models
- 1960s-90s: one of the dominant methods in all of quantitative social science. Now other methods compete with SEMs, but still remains a major part of methodological toolkit for longitudinal and other analyses
- Enjoying a recent renaissance
  - "Directed Acyclic Graphs" (DAGs) and the approach to causality outlined by Judea Pearl (2000) and followers
  - Multilevel Structural Equation Models (MLSEM) in Psychology developed by Muthén, MacKinnon, Preacher and others

Example: Miller and Stokes (1956) Model of Congressional Representation



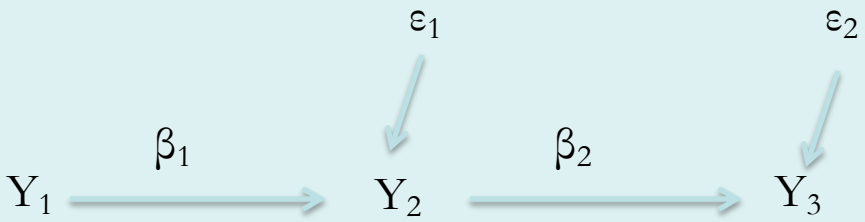
Illustrates features of SEMs:

- *System of Equations*: here modeling three different DVs
- *Exogenous* versus *Endogenous* variables (as opposed to *Independent versus Dependent* variables)
- *Direct* versus *Indirect* Causal Effects
- *Recursive* versus *Non-Recursive* Models
- *Identification* of model parameters: is there enough information in the model to estimate all the coefficients of interest?

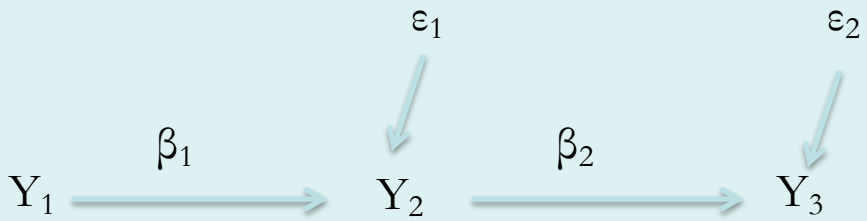
SEM and Longitudinal Analysis: A Natural "Fit"

- Multiple waves of observations means equations for each DV in each time period, so can conceive of longitudinal models as systems of equations over time
- Interest in much longitudinal research is in modeling reciprocal effects between variables, so SEM methods useful
- Similarly, interest in much longitudinal research is in modeling relationships between variables that are "purged" of measurement error, and SEM methods are probably the best available for this purpose
- Interest in much longitudinal research in causal mediation and the separation of *direct* and *indirect* effects, so SEM methods also ideal (e.g. joining groups leads to more civic skills which then feeds into more participation)
- SEM can be used for "Latent Growth Modeling"; this represents one integration of SEM with other longitudinal modeling methods. More general Multilevel SEM models for longitudinal data are also possible, and we'll cover some later in the course

Example: A Simple Longitudinal SEM



- One variable, say *Party Identification* , influencing itself over 3 waves ("autoregressive model")
- Two equations, one predicting PID in wave 2, one predicting PID in wave 3
- Two *endogenous* variables (PID_2 and PID_3), one *exogeneous* variable (PID_1)
- It is a *recursive model*, since no causal "feedback" effects between variables
- β_1 and β_2 are the "structural effects" linking PID in waves 1-2 and 2-3. They are sometimes in longitudinal models called the "stability" effects
- ε_1 and ε_2 are the "disturbance" or "error" terms, unobserved influences and idiosyncratic errors that predict PID in waves 2 and 3
- If we express all variables as deviations from their respective means, there is no need for an intercept term in the Y_2 or Y_3 equations (this can be modified)



How to estimate the structural parameters in this model?

- Run two regressions, one endogenous variable at a time. This is always possible and will give you the *same estimates in simple recursive models* as SEM methods (i.e., whenever models have no measurement error, no correlated disturbances, and no feedback effects).
- Use SEM methods, which uses all the information that is given in the data (i.e., the variances and covariances of all the observed variables), and generates estimates of individual structural coefficients as well as tests of the "fit" of the model as a whole to the data

SEM estimation procedure (non-technical version)

1. Express the variances and covariances of the observed variables in terms of the *unknown* structural effects of the model
2. Manipulate the variances and covariances of the observed variables to arrive at estimates of the model's unknowns. That is, we solve for the *unknowns* in terms of the *known* variances and covariances.
3. If models are "overidentified," i.e., more knowns than unknowns, we can generate *predictions* about the variances and covariances that *must be true* if the model is correct. We then compare the predicted and actual values of the variances and covariances: if they are the same, we say the model "fits" (or "is consistent with") the data; if they are not the same, we "reject" the model as a whole.
4. "Rejected" models then are revised; models that "fit" may or may not be accepted, however. Many different models may produce a good fit to the data. Theory is always relevant!!

SEM analysis is also called "Covariance Structure Analysis" or CSA: we use the observed variances-covariances to estimate structural parameters and we use the estimated structural parameters to generate predicted variances and covariances, which we then compare to the actual ones in assessing how well the model "reproduces" the observed data. Very dialectic!

So with SEM we move from a world where we care only about regression coefficients and the fit (R^2) of a particular equation to a world where we care both about regression coefficients (structural effects), R^2, *and* assessing the overall "fit" of a given model

How do we do this? A Simple Example

$$(1) \quad Y_2 = \beta_1 Y_1 + \varepsilon_1$$

$$(2) \quad Y_3 = \beta_2 Y_2 + \varepsilon_2$$

$$\text{with assumptions } E(\varepsilon_1 \varepsilon_2) = E(Y_1 \varepsilon_1) = E(Y_2 \varepsilon_2) = 0$$

- What are the "knowns" in this system? The observed covariances between PID at time 1 (Y_1), PID at time 2 (Y_2), and PID at time 3 (Y_3)
- If we standardize the variables (i.e. mean of 0, s.d. of 1), we will have a correlation matrix between the 3 variables with values of 1 on the diagonal, and the correlations between PID1-PID2, PID1-PID3 and PID2-PID3 in the off-diagonal cells. There will be 3 distinct pieces of information in the observed correlation matrix

Step 1: Express Variances-Covariances (Correlations) in Terms of Unknown Parameters

$$(1)\quad Y_2 = \beta_1 Y_1 + \varepsilon_1$$

$$(2)\quad Y_3 = \beta_2 Y_2 + \varepsilon_2$$

with assumptions $E(\varepsilon_1 \varepsilon_2) = E(Y_1 \varepsilon_1) = E(Y_2 \varepsilon_2) = 0$

- How do we do this? Some basic *covariance algebra*
- Example: Multiply both sides of equation (1) by Y_1 and take expectations (i.e., the long-run "average" for population)

$$E(Y_1 Y_2) = E(Y_1(\beta_1 Y_1 + \varepsilon_1))$$

$$E(Y_1 Y_2) = E(\beta_1 Y_1 Y_1 + Y_1 \varepsilon_1)$$

$$Cov(Y_1 Y_2) = \beta_1 Var(Y_1 Y_1) + Cov(Y_1 \varepsilon_1)$$

$$Cov(Y_1 Y_2) = \beta_1$$

Rules of Mathematical Expectations

1) $E(a) = a$	"the expectation of a constant is the constant"
2) $E(aX) = aE(X)$	"the expectation of a constant multiplied by a variable is the constant multiplied by the expectation of the variable"
3) $E(XX) = Var(X)$ $E(XY) = Cov(XY)$	"the expectation of a variable multiplied by itself is the variable's variance; the expectation of a variable multiplied by another variable is their covariance"
4) $E(a+X)= a + E(X)$	"the expectation of constant plus a variable is the constant plus the expectation of the variable"
5) $E(X+Y)=E(X)+E(Y)$	"the expectation of the sum of two variables is equal to the sum of the two expectations"

- So our model implies that the observed covariance (correlation here, given standardized variables) between PID at time 1 and 2 is equal to β_1
- Our estimate of the model's structural parameter β_1, then, is going to be $[r(Y_1Y_2)]$
- This is the SEM version of the fact that, in a standardized bivariate regression model, **the regression coefficient is equal to the correlation coefficient between X and Y**
- Carrying out the covariance algebra in this system of equations also generates:
- $Corr(Y_2Y_3)=\beta_2$ and $Corr(Y_1Y_3)=\beta_1\beta_2$

Observed Versus "Implied" Correlations

Observed Correlations

	Y_1	Y_2	Y_3
Y_1	1		
Y_2	$r(Y_1Y_2)$	1	
Y_3	$r(Y_1Y_3)$	$r(Y_2Y_3)$	1

Correlations Implied by the Causal Model

	1		
	β_1	1	
	$\beta_1\beta_2$	β_2	1

Step 2: Solve for unknown structural parameters

- What are unknowns? β_1 and β_2
- Easy to solve for each:

  $\beta_1 = r(Y_1 Y_2)$

  $\beta_2 = r(Y_2 Y_3)$

- These are two bivariate regression equations, so in standardized form, the estimate is equal to the correlation between the IV and the DV
- Multivariate and other more complex models, of course, will have more complex solutions for each unknown, but the principle will be the same

Step 3: Assess Model Fit

- After solving for β_1 and β_2 , can assess "fit" for each equation with tradition measures like R-squared, etc.
- But with SEM, can also test --- under some conditions – whether the model *as a whole* fits the observed data, i.e., do the estimates produced from the algebraic manipulations from previous slide successfully reproduce the observed variances and covariances?
- In this case, we *can* test the model. We see that the model makes the following prediction: $\mathbf{Corr(Y_1Y_3)=\beta_1 * \beta_2}$
- So we solve for β_1 and β_2 and then generate a prediction for $Corr(Y_1Y_3)$. If it is close (given sampling error), then the model is "consistent" with the data; if not, we reject the model and say that it needs to be modified.
- More general test of the model : Y_1 and Y_3 are *conditionally independent, given* Y_2 , or $[rY_1Y_3 \mid Y_2)=0]$. This doesn't require estimation of any β, so allows the relationships to follow any functional/parametric form

"Overidentified" Models

- Assessing model fit is only possible when models are *overidentified*, when there are more knowns than unknowns. In our case we have 2 unknowns (β_1 and β_2) and 3 correlations, so the model is overidentified. Technically we say that the model has 1 "degree of freedom".
- When models are "just-identified", there are an equal number of knowns as unknowns, and the model estimates will reproduce he observed data (correlations) exactly. Also known as a "saturated" model
- When models are overidentified, they will not necessarily reproduce the observed data exactly, thus giving us the "degree(s) of freedom" to test the model.

- Another way to look at it: there are actually *two* ways of arriving at β_2
  - One is our original estimate: $r(Y_2Y_3)$
  - Another is (following covariance algebra multiplying equation for Y_3 by Y_1) :

    $r(Y_1Y_3)/ r(Y_1Y_2)$
- So the test of whether the two estimates are the same, given sampling error, is whether:

  $r(Y_2Y_3)= r(Y_1Y_3)/ r(Y_1Y_2)$
  - If so, the model is consistent with the data. If not, reject the model as specified and need to modify or abandon it

Example:
National Election Panel Study, 2000-2002-2004
Stability of Party Identification

Observed

$$\begin{matrix} 1 & & \\ .872 & 1 & \\ .856 & .884 & 1 \end{matrix}$$

Implied

$$\begin{matrix} 1 & & \\ \beta_1 & 1 & \\ \beta_1\beta_2 & \beta_2 & 1 \end{matrix}$$

Estimates and Model Fit

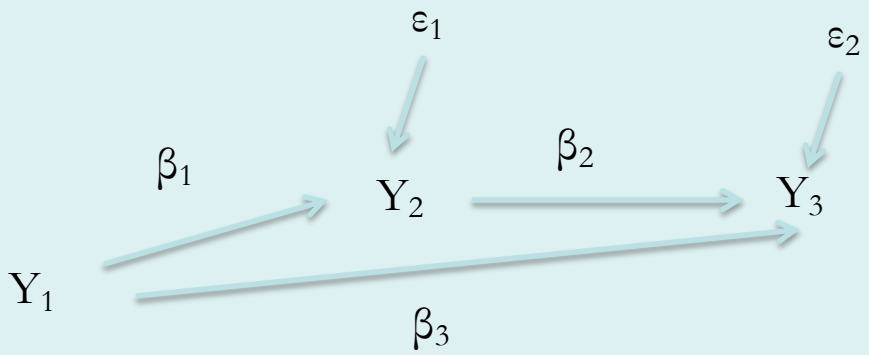
- $\beta_1 = r(Y_1, Y_2) = .872$
- $\beta_2 = r(Y_2, Y_3) = .884$

- Does the model "fit" the data?
- Predicted $r(Y_1\ Y_3)=\beta_1\beta_2=.872*.884=.771$
- But observed $r(Y_1\ Y_3) =.856$, so our model is "wrong" (without considering sampling error)
- Alternatively:

  $\beta_{2=}\ r(Y_1\ Y_3) / r(Y_1\ Y_2) =.856/.872=.98$

  A much different from the earlier estimate of .884
- What is wrong with the model? No direct effect from $PID_{2000} \rightarrow PID_{2004}$ (probably, but could be other omissions)

Alternative Model: Include direct effect of PID_{2000} on PID_{2004}



Observed Correlations

	Y_1	Y_2	Y_3
Y_1	1		
Y_2	$r(Y_1Y_2)$	1	
Y_3	$r(Y_1Y_3)$	$r(Y_2Y_3)$	1

Model Implied Correlations

	Y_1	Y_2	Y_3
Y_1	1		
Y_2	β_1	1	
Y_3	$\beta_1\beta_2+\beta_3$	$\beta_2+\beta_1\beta_3$	1

- This model has three unknowns (β_1, β_2 , and β_3) and three knowns
- It is "*saturated*" or "*just identified*", with 0 degrees of freedom. In recursive models it means that no structural effects are left out of the model
- **As noted earlier, you can't test these models in terms of fit because the model will *necessarily* reproduce the observed variances and covariances exactly**
- You can see that we are going to manipulate all of the correlations to obtain estimates of the three unknowns, and so we have no "excess" correlations with which we can test the fit of the model (like we did in the last model with Corr(Y_1Y_3)
- We can say that the restriction that $\beta_3=0$ in the previous model is what allowed us to test it – it is the "overidentifying restriction" of that model

Maximum Likelihood Estimation of SEM Models

- Basic idea: Find the parameters that generate the implied population variance-covariance matrix ($\hat{\Sigma}$) of the observed variables that comes as close as possible to the actual variance-covariance matrix **S**. Those parameters then *maximized the likelihood of having observed the variance covariance matrix we did observe*
- Test significance of individual coefficient estimates
- Use summary statistics to evaluate the fit of the model as a whole:
- Modify and compare alternative models, especially models "nested" within one another, i.e., with same variables and structure aside from one or more constraints on parameter values in a "reduced" versus a "full" model

How to Estimate SEM Models via ML?

- Two matrices:

S = Observed variances and covariances in our sample, which was randomly drawn from a population with variances-covariances Σ

$\hat{\Sigma}$= Implied variances-covariances based on unknown model parameters (in our case β and the variances-covariances of the ε errors; if the variables are not standardized then there are also variances of the individual variables to take into account, i.e., the diagonal elements of $\hat{\Sigma}$ which are all 1 in a standardized model)

- **THE METHOD OF MAXIMUM LIKELIHOOD FINDS THE VALUES OF THE UNKNOWN MODEL PARAMETERS THAT, TAKEN TOGETHER, MINIMIZE THE DIFFERENCE BETWEEN THE VALUES OF S AND $\hat{\Sigma}$**

Implied Variance-Covariance Matrix: 3 Wave Autoregressive Model (with Unstandardized Variables)

$$\hat{\Sigma} = \begin{bmatrix} \beta_2^2\beta_1^2\phi_1 + \beta_2^2\varphi_1 + \varphi_2 & & \\ \beta_2\beta_1^2\phi_1 + \beta_2\varphi_{11} & \beta_1^2\phi_1 + \varphi_1 & \\ \beta_2\beta_1\phi_{11} & \beta_1\phi_1 & \phi_1 \end{bmatrix}$$

Y_3 Y_2 Y_1

Notes:

- Matrices usually ordered from latest to earliest waves and from endogenous to exogenous variables
- Variances of endogenous variables' error terms (ε) denoted with Greek "psi" ψ
- Variances of exogenous variables denoted with Greek "phi" φ
- Effects of exogenous variables onto endogenous variables typically denoted with Greek "gamma"γ; Y_1 in this example is exogenous so (technically) β_1 is really γ_1
- Stata and other SEM programs' output will have "Gamma" and "Beta" matrices to distinguish these effects

- ML Estimation: What values of $\beta_2, \beta_1, \phi_1, \varphi_1, \text{and } \varphi_2$ will produce the $\hat{\Sigma}$ that is "closest" to **S**?
- When models are "just-identified" or "saturated," the algebraic manipulation of **S** will produce one unique estimate of each of the unknown parameters, and hence the implied $\hat{\Sigma}$ will reproduce **S** ***exactly***
- But when models are "overidentified," there will be several, perhaps many different ways of manipulating to produce the unknown parameters, and hence there is a need for some estimation procedure to "combine" or find the "best" estimates.

- Technically:
- ML estimates are those that *minimize* a fit function (call it F) that expresses the deviation between $\hat{\Sigma}$ and **S** as:

$$F = \log\left|\hat{\Sigma}\right| + tr\left(\frac{S}{\hat{\Sigma}}\right) - \log|S| - t$$

where the first and third terms are the "determinants" of $\hat{\Sigma}$ and **S**, respectively, *tr* is the "trace" of a matrix, and *t* is the total number of variables in the model. ML estimates are produced via iterative processes, where initial values are adjusted, recalibrated, etc. until no "better" estimates can be produced.

- If you are interested in learning about different algorithms used in ML estimation of this sort, see Kenneth Bollen, *Structural Equation Models for Latent Variables* (John Wiley and Sons, 1989), pp.131-144.
- When $\hat{\Sigma}$ and **S** are equal, F will be 0, so the goal is for F to be as small as possible. It will be 0 only when the model perfectly reproduces S, e.g. with a saturated model.

Results From Our Example

- For Matrix A with elements $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

- Our $\hat{\Sigma}$ (from STATA post-estimation command "estat framework, fitted")

	observed		
Sigma	pid2004	pid2002	pid2000
observed			
pid2004	**5.42207**		
pid2002	**4.468679**	**4.746776**	
pid2000	**3.865864**	**4.106447**	**4.675995**

- $\det(\hat{\Sigma})=6.49$

- Our **S** (from Stata command "corr pid2004 pid2002 pid2000, cov")

	pid2004	pid2002	pid2000
pid2004	**5.42877**		
pid2002	**4.4742**	**4.75264**	
pid2000	**4.30129**	**4.11152**	**4.68177**

- det(**S**)=5.63
  (http://ncalculators.com/matrix/matrix-determinant-calculator.htm)
- The trace of a matrix is simply the sum of the diagonal elements. In the case of $S/\hat{\Sigma}$ we would divide the diagonal elements of S by the diagonal elements of $\hat{\Sigma}$ and sum the three terms. In our case, since we reproduce all of the diagonal variances exactly with our model, we have values of 1 on the 3 diagonals, so the trace of this matrix is 3. Then we subtract *t*, which is also 3 for the number of variables, so we see that the second and fourth terms of the equation for ***F*** cancel out in this case.
- So: F = log (6.49) – log (5.63) = 1.87-1.728= .142

Testing Model Fit

- H_0: The covariance matrix in the population is equal to $\hat{\Sigma}$, the covariance matrix implied by our model. (Informally: our model "fits" the data, taking sampling error into account)
- H_A: The covariance matrix in the population is equal to "any arbitrary matrix", and if it is arbitrary, we will pick **S** when maximizing the likelihood function. (Informally: no "model" at all is needed to estimate, only our sample variances and covariances).
- So under each of these two hypotheses, we can generate a (log) Likelihood function that summarizes the overall likelihood that the specified model holds in the population. We can then test H_0 by comparing the *ratio* of the two (log) Likelihoods.
- The ratio of the two log Likelihoods (the "Likelihood Ratio Chi-Square,") is equal to (N-1)*F, and is distributed as a χ^2 statistics with degrees of freedom equal to $((t(t+1)/2) - k)$, where t is the number of variables, and k is the number of unknown parameters in the model. Here we have 6 knowns and 5 unknowns, so df=1.

- This means that the Likelihood Ratio chi-square (call it L^2) can be used as a statistical test of the null hypothesis that the specified model holds in the population.
- Note that this is fundamentally different than other null hypotheses you have encountered, where the null represents *NO* relationship or no effect from one variable to another. Here our null is that the model fits the data, and we therefore DO NOT want to reject it. So large Likelihood ratio chi-square values means we reject the null and our model does not fit the data, or is "inconsistent" with the data.
- Our example: (N-1)*F, the "Likelihood Ratio Chi-square." Here it is 809*.142 = 115.0 (close enough to the 117.9 from STATA output). With 1 degree of freedom, the statistical significance of this figure is extremely high, that is, it was very unlikely to have come about by chance sampling error. So we can say that the deviations between $\hat{\Sigma}$ and $\mathbf{S}$ are "statistically significant", that is, **there is a significant difference between our model's $\hat{\Sigma}$ and the true population covariance matrix Σ.**
- **This means our model is "inconsistent with the data."**

Testing Alternative Models

- The same logic can also be used to test the statistical significance of any *improvements in model fit* associated with relaxing model constraints, that is, in comparing the fit of two different models.
- Suppose we relax the constraint that $\beta_3 = 0$. Whenever one model can be arrived at by imposing one or more constraints on another model, we say that the model is "nested" in the other model. In such cases, the *difference in* χ^2 between the two models itself follows as χ^2 distribution, and the improvement in fit can be tested according to whether the difference in χ^2 is statistically significant, given the difference in degrees of freedom (or the number of imposed constraints) between the two models.
- In this case:

  "Constrained" Model (where $\beta_3 = 0$: $\chi^2 = 117.9$. w/ 1 df

  "Unconstrained" Model (where β_3 is a free parameter): $\chi^2 = 0$ w/ 0 df

  χ^2 Difference (Constrained – Unconstrained) = 117.9 df Difference=1

  χ^2 Difference is **statistically significant** and therefore the Unconstrained model represents a *statistically significant improvement* over the Constrained model in terms of reproducing **S (or Σ)**.

Notes on Model Testing and Assessing Goodness of Fit

1. In this case the "unconstrained model" is a saturated model but this need not be the case. An "unconstrained model" is *any* model that relaxes one or more constraints on model parameters, and the difference in χ^2 between the two models provides a test of the significance of the improvement in fit from relaxing the constraints.
2. It is also the case that this is the only kind of test that you can do with saturated models, that is, compare them to other models in terms of fit. You cannot conclude anything about a saturated model *in itself* because the fit is by definition perfect.
3. An insignificant model χ^2 or L^2 does not mean that your model is "correct" or the "true" model. It is just "consistent with the data". There could be many models that reproduce the observed data as well as yours. **So a significant model χ^2 or L^2 means that the model is inconsistent with the data and can be rejected, but an insignificant χ^2 or L^2 does not mean the model is accepted as "the truth".**

4. Since the Likelihood Ratio χ^2 (L^2) depends on N and F, you can in many instances increase the size of χ^2 simply by increasing the number of cases in the analysis. This means that the same degree of fit will "look worse" with a larger N, a somewhat perverse situation in social science, as we usually want to maximize the number of observations.

5. Hence methodologists have come up with various ways of correcting L^2. Sometimes the correction is done by dividing L^2 by the degrees of freedom in the model, but no one knows how to interpret L^2/df with any specificity. It used to be said that "values approaching 2-3 are beginning to be reasonable", but this is no longer accepted.

One "chi-square correction" measure of fit that is very popular in the SEM literature is the "Root Mean Square Error of Approximation" (RMSEA),:

$$RMSEA = \sqrt{\frac{\chi^2_m - df_m}{(N-1)df_m}}$$

which in this case is: square root of 116.9/809=.38

It is desirable for this value NOT TO EXCEED .08 for models to be acceptable!!!

6. Other measures of goodness-of-fit involve comparing the model chi-square in various ways to a very crude "baseline model" which has *no parameters* whatsoever and no covariances between the observed variables – it just has variances of variables that are completely unrelated in the population. (This is about as "baseline" as you can get!)

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(1)	**117.941**	model vs. saturated
p > chi2	**0.000**	
chi2_bs(3)	**2484.410**	baseline vs. saturated
p > chi2	**0.000**	

- Note the baseline model has 3 df since there are 3 observed variables (i.e., 3 variances are estimated and that is it!)

- We can calculate the extent to which our model improved on the baseline model via a proportional reduction in the chi-square value, or the Comparative Fit Index (CFI):

$$CFI = \frac{\chi_b^2 - \chi_m^2}{\chi_b^2}$$

- So the CFI is (2484.4-117.9)/2484.4=.953
- Tucker and Lewis (1973) suggest correcting the CFI to take into account the relative degrees of freedom in the two models, so that near-saturated models are penalized to a greater extent:

$$TLI = \frac{\chi_b^2 / df_b - \chi_m^2 / df_m}{\chi_b^2 / df_b - 1}$$

which equals (828.1-117.9)/828.1=.857

- CFI and TLI should be *greater than .90 and close to .95 if possible!!!!*

7. Last measures of Goodness of Fit

A. The "Average" residual, or square root of the total squared (standardized) deviations between the observed and predicted covariances. This is the "**Standardized Root Mean Square Residual" (SRMSR)** in STATA, and should not be above .08 (some say .05) for model acceptability. Here we obtain .028, so not bad.

B. "Information"-based measures. These are the only measures that do not depend on models being nested in one another. Akaike's Information Criteria (AIC) is:

$$AIC = -2\ln L_m + 2 * k$$

where L is the model's log-likelihood (NOT CHI-SQUARE) and k is the number of estimated parameters. Smaller numbers are better, so the 2*k term is a penalty for the expression as more parameters are estimated. Used very often to compare models with different variables, structures, etc., and we will see this in later sections of the course as well.

Here AIC=(-2*-4204.89)+2*5=8420

All STATA Goodness of Fit measures from "estat gof, sta(all)"

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(1)	**117.941**	model vs. saturated
p > chi2	**0.000**	
chi2_bs(3)	**2484.410**	baseline vs. saturated
p > chi2	**0.000**	
Population error		
RMSEA	**0.380**	Root mean squared error of approximation
90% CI, lower bound	**0.324**	
upper bound	**0.439**	
pclose	**0.000**	Probability RMSEA <= 0.05
Information criteria		
AIC	**8421.784**	Akaike's information criterion
BIC	**8449.966**	Bayesian information criterion
Baseline comparison		
CFI	**0.953**	Comparative fit index
TLI	**0.859**	Tucker-Lewis index
Size of residuals		
SRMR	**0.028**	Standardized root mean squared residual
CD	**0.760**	Coefficient of determination